AQL 2021

# The 14th International Conference on Advances in Quantitative Laryngology, Voice and Speech Research

# CONFERENCE PROCEEDINGS

2021

**14TH INTERNATIONAL CONFERENCE**

**ADVANCES IN QUANTITATIVE LARYNGOLOGY,**

**VOICE AND SPEECH RESEARCH (AQL)**

**JUNE 7-10, 2021**

**BOGOTÁ, COLOMBIA (ONLINE)**

# CONFERENCE PROCEEDINGS

## EDITED BY

LADY CATHERINE CANTOR CUTIVA

**ISSN:** 2745-2751 (En línea)

**Key-title:** Proceeding 14th International Conference on Advances in Quantitative Laryngology, Voice and Speech Research (Online)
**Title proper:** Proceeding 14th International Conference on Advances in Quantitative Laryngology, Voice and Speech Research
**Abbreviated key-title:** Proc. 14th Int. Conf. Adv. Quant. Laryng. Voice and Speech Res.
**Other variant title:** AQL 2021
**Original alphabet of title:** Times New Roman
**Subject:** Quantitative Laryngology, Voice and Speech Research
**Publisher:** AQL Local Organizing Committee
**Dates of publication:** 2021
**Frequency:** Every 2 years
**Language:** English
**Country:** Colombia
**Medium:** Online
**Last modification date:** June 23, 2021
**ISSN Center responsible of the record:** Biblioteca Nacional de Colombia
**URL:** https://aql2021.sciencesconf.org/data/pages/210624_BOOK_OF_ABSTRACTS_complete.pdf

AMA style reference:

Cantor-Cutiva, L.C, ed. *Proceedings of the 14th International Conference on Advances in Quantitative Laryngology, Voice and Speech Research*. Bogotá, Colombia. 2021;14.

# AQL 2021 LOCAL ORGANIZING AND SCIENTIFIC COMMITTEE

| | |
|---|---|
| **Lady Catherine Cantor Cutiva** (Chair) | Universidad Nacional de Colombia<br>Universidad Manuela Beltrán, Colombia |
| **Luisa Fernanda Ángel Gordillo** | Universidad Nacional de Colombia<br>Hospital Nacional Universitario, Colombia |
| **Elisa Landázuri Cocinero** | Universidad Manuela Beltrán, Colombia |
| **Hédrick Yoseft Robles Vega** | Universidad Simón Bolívar, Colombia |
| **Pasquale Bottalico** | University of Illinois at Urbana-Champaign, USA |
| **Maryam Naghibolhosseini** | Michigan State University, USA |
| **Russell Banks** | Massachusetts General Hospital, USA |
| **Carlos Alberto Calvache** | Corporación Universitaria Iberoamericana, Colombia |

# AQL 2021 SPONSORS

**National Institute on Deafness and Other Communication Disorders, National Institutes of Health**

**Universidad Manuela Beltrán**

**Vocology Center**

**Revista de Investigación e Innovación en Ciencias de la Salud (RIICS)**

**Colegio Colombiano de Fonoaudiólogos (CCF)**

# List of Content

# COMPARISON OF ACOUSTIC SIGNAL QUALITY WITH RESPECT TO GLOTTAL CLOSURE BETWEEN OLD AND YOUNG EX VIVO SHEEP LARYNGES

**Bernhard Jakubaß[1], Markus Gugatschka[2], Claus Gerstenberger[2], Marion Semmler[1], Michael Döllinger[1]**

[1] Division of Phoniatrics and Pediatric Audiology, Department of Otorhinolaryngology, University Hospital Erlangen, FAU Erlangen-Nürnberg, Germany

[2] Division of Phoniatrics, ENT University Hospital Graz, Medical University Graz, Austria

***Keywords*: Presbyphonia, Larynx, High-speed imaging, Ex-vivo, Glottal closure**

## INTRODUCTION

With age, the quality of the voice decreases due to the loss of muscle flexibility and the degeneration of nerves, called presbyphonia. The presence of presbyphonia can influence the life quality of the elderly, as the confidence to speak decreases, which can lead to social isolation. In preparation for the evaluation of functional electrical stimulation (FES) as a therapy method of presbyphonia, our aim for this study was to establish a data set of untreated old and young sheep ex-vivo larynges. In addition to previously published subglottal pressure and audio data [1, 2], we now analyzed the high-speed imaging (HSI) data. For this study, we compare the acoustic quality between young and old sheep as function of glottal gap sizes.

## METHODS

After harvesting, the larynges of twelve young sheep and six old sheep were quick-frozen in liquid nitrogen and stored in a freezer (-80 °C) to prevent degeneration. For the experiments, the larynges were thawed and then mounted on an artificial trachea. The epiglottis and parts of the thyroid cartilage were removed for a better view of the vocal folds. The vocal folds were adducted towards the phonation position by metal rods. Three different weights (20 g, 40 g, and 60 g) were mounted to the thyroid cartilage to elongate the vocal folds. For each weight, 16 runs with different airflow levels were performed. First, the flow was increased gradually until the phonation onset. From there on, the flow was increased in steps of 2.5 slm or 5 slm for 15 times. The experiments were captured by a subglottal pressure sensor as well as a supraglottal microphone and a high-speed camera. These three signals were triggered and captured synchronously by a LabVIEW script. The subglottal pressure and the audio signal were recorded for 1 s at a sampling rate of 96 kHz, while the HSI recording was performed with 4000 frames per second for 0.5 s, due to the limitation of memory capacity of the camera. The glottal area was segmented from the HSI data by using our *Glottis Analysis Tools* software (GAT). HSI based Glottal Gap Index (GGI) and Cepstral Peak Prominence (CPP), as quality measure for the acoustic and subglottal pressure, were computed using GAT. Further analysis was performed using MATLAB and SPSS.

## RESULTS

Based on the GGI derived from the HSI recordings, we will split the data set into three groups separated for both age groups: (1) complete closure (GGI = [0; 0.01]), (2) partial closure (GGI = ]0.01; 0.4[), and (3) no closure (GGI = [0.4; 1]). For these three closure characteristics, we will present and discuss acoustic quality, based on CPP, for the different closure types and age groups. First results show a decrease of CPP values with increasing glottal gap (i.e., GGI increases) for young as well as old sheep.

## CONCLUSION

With the evaluation of these data, we continue to establish a control data set for the evaluation of subsequent FES experiments as potential therapy method for presbyphonia in future.

## ACKNOWLEDGMENTS

The authors acknowledge support from the German Research Foundation (DFG) under DO 1247/12-1 (no. 409543779) and the Austrian Research Council (FWF) under no. I 3997-B27.

## REFERENCES

[1] Gerstenberger, Claus; Döllinger, Michael; Kniesburges, Stefan; Bubalo, Vladimir; Karbiener, Michael; Schlager, Hansjörg; Sadeghi, Hossein; Wendler, Olaf; Gugatschka, Markus (2018): Phonation Analysis Combined with 3D Reconstruction of the Thyroarytenoid Muscle in Aged Ovine Ex Vivo Larynx Models. In: *Journal of voice* 32 (5), S. 517–524. DOI: 10.1016/j.jvoice.2017.08.016.

[2] Döllinger, Michael; Wendler, Olaf; Gerstenberger, Claus; Grossmann, Tanja; Semmler, Marion; Sadeghi, Hossein; Gugatschka, Markus (2019): Juvenile Ovine Ex Vivo Larynges: Phonatory, Histologic, and Micro CT based Anatomic Analyses.

*The 14th International Conference on Advances in Quantitative Laryngology, Voice and Speech Research (AQL)*

In: *BioMed Research International* 2019. DOI: 10.1155/2019/6932047

# EFFECTS OF AGE, SEX, AND PARKINSON'S DISEASE ON KINEMATIC AND ACOUSTIC FEATURES OF PHONATORY OFFSET

**Jennifer M. Vojtech[1,2*], Cara E. Stepp[1,2,3]**

[1] Department of Biomedical Engineering, Boston University, Boston, MA 02215, USA
[2] Department of Speech, Language, and Hearing Sciences, Boston University, Boston, MA 02215, USA
[3] Department of Otolaryngology—Head and Neck Surgery, Boston University School of Medicine, Boston, MA 02118, USA
[*]Current Affiliation: Delsys, Inc. and Altec, Inc., Natick, MA 01760, USA; jvojtech@delsys.com

***Keywords*: Voice; Vocal Fold Abduction; High-Speed Videoendoscopy; Relative Fundamental Frequency**

## INTRODUCTION

Laryngeal muscle tension and vocal fold abductory kinematics play key roles in the regulation of intervocalic offsets. Yet the contribution of abduction to measures of relative fundamental frequency (RFF)—an acoustic estimate of the degree of laryngeal tension—has not been physiologically assessed. Thus, the aim of this study was to examine the relationship between kinematic and acoustic features of phonatory offset in two experiments. Experiment 1 assessed the relationship between vocal fold abduction and RFF at voicing offset as a function of speaker age and sex. We hypothesized that abductory kinematics would be significantly predictive of RFF. Experiment 2 quantified the effects of PD on vocal fold abduction and RFF. We hypothesized that measures of abduction would not significantly differ between speakers with and without PD, but that speakers with PD would exhibit significantly lower RFF values at voicing offset.

## METHODS

### *Participants*

In Experiment 1, 50 adults with typical voices (25 cisgender females, 25 cisgender males) aged 18–83 years were enrolled in the study. In Experiment 2, 20 adults with idiopathic PD (6 cisgender females, 14 cisgender males) aged 50–75 years and 20 age- and sex-matched controls aged 47–81 years were enrolled in the study.

### *Instrumentation and Measurement*

In both experiments, simultaneous acoustic and high-speed videoendoscopic recordings were acquired from participants producing the utterance /ifi/. Vocal fold abduction was characterized for each /ifi/ production via measures of abduction duration (AD) and glottic angle at voicing offset (GA). Estimates of AD and GA were extracted from the laryngoscopic images. RFF was calculated from the acoustic signal using a semi-automated algorithm [1].

### *Analysis*

In Experiment 1, the relationship of RFF with AD, GA, age, and sex was quantified via an analysis of covariance (ANCOVA). In Experiment 2, three one-way analyses of variance were constructed to assess the effect of speaker group on RFF, AD, and GA.

## RESULTS AND DISCUSSION

In Experiment 1, only GA was a significant factor ($p$ = .019, $\eta_p^2$ = 0.12) in the model for RFF (**Table 1**). This suggests that RFF was related to abduction during devoicing, but that this relationship was not significantly impacted by speaker age or sex.

**Table 1: Results of ANCOVA examining the effects of speaker age, speaker sex, AD, and GA on RFF.**

| Effect | df | $F$ | $p$ | $\eta_p^2$ |
|---|---|---|---|---|
| **Age** | 1 | 2.38 | .130 | 0.05 |
| **Sex** | 1 | 3.01 | .090 | 0.06 |
| **AD** | 1 | 0.02 | .895 | 0 |
| **GA** | 1 | 5.89 | .019 | 0.12 |

In Experiment 2, speaker group exhibited a significant effect on RFF ($p$ = .021, $\eta_p^2$ = 0.13) and AD ($p$ = .034, $\eta_p^2$ = 0.11), but not on GA ($p$ = .476, $\eta_p^2$ = 0.01). These findings indicate that kinematic and acoustic measures of phonatory offset were both significantly impacted by PD. Overall, these findings support vocal fold abductory patterns as a mechanism of RFF, wherein changes in RFF during devoicing may be captured in part via glottic angle.

## CONCLUSIONS

These results extend our understanding of the physiological underpinnings of RFF. Future work is needed to investigate the differential contributions of laryngeal tension and vocal fold abduction mechanisms to RFF at intervocalic offsets.

## ACKNOWLEDGMENTS

This work was supported by NSF Grant #1247312 (J.M.V.) and NIH Grant #DC015570 (C.E.S.).

## REFERENCES

[1] Vojtech JM, et al. Refining algorithmic estimation of relative fundamental frequency: Accounting for sample characteristics and fundamental frequency estimation method. *J Acoust Soc Am* 2019; 146:3184–3202.

# INVESTIGATING BLUNT FORCE TRAUMA TO THE LARNYX: THE ROLE OF VERTICAL MISALIGNMENT AND VF SCARRING

**Molly E. Stewart[1], Byron D. Erath[1]**

[1] Department of Mechanical and Aeronautical Engineering, Clarkson University, Potsdam, NY, USA

***Keywords*: Blunt force laryngeal trauma, Vocal fold asymmetry, Voice restoration, Vocal fold scarring**

## INTRODUCTION

Blunt force trauma (BFT) to the larynx is a potentially life-threatening injury that can result from motor vehicle collisions, sports-related injuries, strangulation, and clothesline-type injuries [1]. Treatment following BFT focuses on airway management as the top priority, leaving vocal outcomes as a secondary consideration. The most common injuries following BFT are displaced laryngeal cartilage fractures, vocal fold (VF) scarring, and paralysis of at least one VF [2]. Both displaced cartilage fractures and VF paralysis can cause vertical misalignment between the VFs. It is unclear how much vertical VF misalignment can be tolerated before voice quality degrades significantly, and how scarring influences this relationship. Hence, the objective was to assess how objective aerodynamic and acoustic measures of voice quality vary as a function of vertical displacement between the opposing VFs.

## METHODS

Synthetic, self-oscillating silicone VF models were used in a physiologically-representative flow facility of the lungs, trachea, and subglottal tract. Flow rate, subglottal pressure, VF kinematics, and radiated sound pressure level were acquired.

Vertical displacements in the inferior-superior direction were introduced to one VF in 1.0 mm increments using 3D printed shims. The amount of displacement in the superior direction was defined by the variable $d$ and is presented as a ratio relative to the inferior-superior length of the medial surface of the VF models, $l_{VF}$. When $d > l_{VF}$, the VFs no longer contact during oscillation.

VF scarring was modeled by increasing the stiffness of the cover layer of the VF models. It has been shown that scarred VF tissue can be up to three times stiffer than normal VF tissue [3]. The different cover layer stiffnesses studied are displayed in table 1.

**Table 2: Modulus of Elasticity of the Cover layer for the three different types of models used.**

| | Normal Model | Model 1 | Model 2 |
|---|---|---|---|
| Cover layer | 1.12 kPa | 3.43 kPa | 10.34 kPa |

The combined effect of vertical displacement with VF scarring (i.e., varying cover stiffness) was investigated for the scarred VF being displaced superiorly to the normal VF, as well as the opposite case, with the normal VF being displaced superiorly relative to the scarred VF.

Aerodynamic, kinematic, and acoustic parameters of VF function were measured. Aerodynamic measures included onset pressure, subglottal pressure, and flow rate. Kinematic parameters extracted from high-speed videos included amplitude and phase differences of oscillation. Acoustic parameters included SPL, frequency, H1-H2, jitter, shimmer, and cepstral peak prominence (CPP).

## RESULTS AND DISCUSSION

Significant findings indicate that if the inferior-superior VF misalignment exceeds the inferior-superior medial length of the VF, both acoustic and kinematic measures become pathological, indicative of severely-degraded vocal quality. In particular, the target SPL is no longer attainable, jitter and shimmer values both surpass the threshold to identify pathological voices, phase shift between VFs increases, and CPP decreases. The introduction of stiffer models hastens these affects.

## CONCLUSION

Improved vocal outcomes are expected when VF contact is maintained during phonation (i.e., vertical displacement does not surpass VF medial length) and when stiffness asymmetries are minimized.

## ACKNOWLEDGMENTS

This work was funded in part by the National Institutes of Health (NIH) National Institute on Deafness and other Communication Disorders Grant P50 DC015446.

## REFERENCES

[1] Butler, A.P., O'Rourke, A.K., Wood, B.P., Porubsky, E.S. Acute external laryngeal trauma: experience with 112 patients. *Annals of Otology, Rhinology & Laryngology* 2005;114(5): 361-368.

[2] Heman-Ackah, Y.D., Sataloff, R.T. Blunt trauma to the larynx and trachea: considerations for the professional voice user. *J Sing* 2002:59, 41-7.

[3] Thibeault, S.L., Gray, S.D., Bless, D.M., Chan, R.W., Ford, C.N. Histologic and rheologic characterization of vocal fold scarring. *J Voice* 2002:16(1), 96-104.

# DISCOVERING UNDERLYING PHYSICAL PARAMETERS FROM DAILY PHONOTRAUMA INDEX DISTRIBUTIONS USING MONTE CARLO SIMULATIONS OF A LOW-DIMENSIONAL VOICE PRODUCTION MODEL

**Parra Jesús A.[1], Ibarra Emiro[1], Alzamendi Gabriel A.[1,2], Cortés Juan P.[3], Zañartu Matías[1]**

[1] Department of Electronic Engineering, Federico Santa María Technical University, Valparaíso, Chile
[2] Institute for Research and Development on Bioengineering and Bioinformatics, Oro Verde, Entre Ríos, Argentina
[3] Center for Laryngeal Surgery and Voice Rehabilitation Laboratory, Massachusetts General Hospital, MA, United States

***Keywords*: Phonation Model; Hyperfunction Classification; Monte Carlo Simulations; Population Statistics.**

## INTRODUCTION

The Daily phonotrauma index (DPI) is a statistical classifier for ambulatory voice monitoring of phonotraumatic vocal hyperfunction (PVH) patients, by means of a logistic regression framework [1]. While the clinical validity of this classifier has been initially demonstrated, the physical mechanisms that underlie it remain unclear. A model-based simulation framework is proposed to elucidate the physical underpinnings of the DPI and to provide further insights into what we currently classify as PVH in the ambulatory data.

## METHODS

We posit that a quasi-steady representation using a collection of sustained phonatory gestures is sufficient to capture long-term ambulatory group behaviors. Therefore, we incorporate a fully interactive Triangular Body-Cover vocal fold Model into extensive Monte Carlo simulations covering a wide range of model parameters to mimic joint experimental distributions of selected ambulatory measures (e.g., SPL, $H_1$-$H_2$, fo, etc.) for PVH patient and control populations. The resulting distributions provide an inverse mapping relating DPI space to model parameters, such as subglottal pressure and muscle activation.

### *Participants*

The DPI space is presented in [1], it is a scatterplot of SPL skewness and H1-H2 standard deviation, each point represents one person in this study. these statistics are used to transform the mean subject distribution into specific person distribution for SPL and $H_1$-$H_2$, these distributions were mimic using Monte Carlo simulations of the phonation model.

### *Analysis*

The TBCM is a model of sustained vowel phonation, its inputs are pulmonary pressure, muscle activations and shape of the vocal tract, and as an output for this case we have glottal flow and output pressure, from which we compute the characteristics present in the space of DPI. Therefore, by mimicking the distributions of the outputs using Monte Carlo simulations, distributions of the inputs are obtained.

## RESULTS AND DISCUSSION

The resulting distribution for the physical parameters in the DPI space illustrates that the PVH zone in this classifier is associated with increased contact pressure, subglottal pressure, and LCA muscle activation as well as lower CT activation when compared to the healthy control zone.

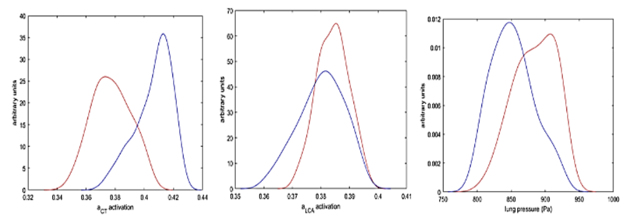


Figure 1. Statistical distribution of mean behavior for parameters from two zones in the DPI space: patient population (red), control population (blue). CT activation (left), LCA activation (center), Lung pressure (right).

## CONCLUSION

Current results suggest that PVH patients operate their voices with high effort compensatory mechanisms throughout the day, which is likely to maintain the PVH vicious cycle. Further investigation is needed to assess the effects of therapy and biofeedback with this approach.

## 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] H. Van Stan, D. D. Mehta, A. J. Ortiz, J. A. Burns, L. E. Toles, K. L. Marks, M. Vangel, T. Hron, S. Zeitels, and R. E. Hillman, Differences in Weeklong Ambulatory Vocal Behavior Between Female Patients With Phonotraumatic Lesions and Matched Controls," JSLHR, vol. 63, no. 2, pp. 372-384, 2020.

# MODELLING OF AMPLITUDE MODULATED VOCAL FRY GLOTTAL AREA WAVEFORMS USING AN ANALYSIS-BY-SYNTHESIS APPROACH

**Vinod Devaraj[1,2], Philipp Aichinger[1]**

[1] Department of Otorhinolaryngology, Division of Phoniatrics-Logopedics, Medical University of Vienna, Austria
[2] Department of Signal Processing and Speech Communication Laboratory, Technische Universität Graz, Austria

***Keywords*: Voice Quality; High-Speed Videoendoscopy; Glottal Area Waveforms; Modelling**

## INTRODUCTION

Characterization of voice quality is important for the diagnosis of voice disorders. Vocal fry is a voice quality which is traditionally characterized by a low frequency and a long-closed phase of the glottis, which gives an auditory impression of "a stick being run along a railing", "popping of corn" or "cooking of food on a pan" [1, 2]. However, we also observed amplitude modulated vocal fry glottal area waveforms (GAWs) without long closed phases. Several other studies have investigated the vibration patterns of vocal fry which include the number of opening and closing phases of the vocal folds in a single amplitude modulator cycle, and the duration of the closed phase [1, 2]. These studies already provided evidence for the existence of multiple pulses in a single modulator cycle.

In the past, acoustic features were mainly used for objective detection of vocal fry. Presence of vocal fry segments in speech utterances were detected based on the autocorrelation properties of the audio signals [4]. In [5], audio features like inter-frame periodicity, inter-pulse similarity, peak fall, and peak rise, H2-H1, i.e., the difference in amplitudes of the first two harmonics, F0 contours in each frame and peak prominence were used for vocal fry detection. A Fourier spectrum analysis approach of the audio signals was also proposed for distinguishing vocal fry segments from diplophonic voice [6]. Though these methods detect vocal fry or creaky segments, they do not allow a detailed study of voice production. In this study, we distinguish amplitude modulated vocal fry (positive group) and euphonic (negative group) voice qualities using GAWs as input data. With GAWs, the cause effect relationship between voice production and perception could be studied in addition.

## METHODS

### *Data collection*

Videos of the vocal folds during sustained phonation were obtained by means of a laryngeal high-speed camera with a frame rate of 4000 frames per second. Audio files were recorded simultaneously. The voice samples were annotated by three listeners regarding presence or absence of vocal fry based on the perception of audio signals. The corresponding GAWs were extracted from high-speed videos. Seven GAWs annotated as vocal fry contained amplitude modulations without long phases. They are used in this study as positive data. A group of eight euphonic GAWs is used as a negative data. In addition to natural GAWs, 1200 synthetic GAWs are used, of which 300 belong to the euphonic group and the remaining 900 are vocal fry GAWs.

### *Modelling GAWs*

We model input GAWs using an analysis-by-synthesis approach [7, 8]. First, fundamental frequency $f_o$ of an input GAW is extracted by a hidden Markov model (HMM) combined with repetitive execution of a Viterbi algorithm. An unmodulated quasi-unit pulse train is generated by an oscillator driven by the extracted $f_o$ track. The pulse locations of this pulse train approximate the time instants of the maxima of the input GAW. An additional pulse train for indicating the locations of the minima of the input GAW is obtained by phase shifting the train's phase by $180^o$. The instantaneous phase and amplitude of the maxima and minima are extracted from the quasi-unit pulse trains.

Pulse shapes are obtained by cross-correlating the quasi-unit pulse with each block of the input GAW of length 32ms obtained using a Hanning window with a 50 percent overlap. The pulse shapes $r_l$ are transformed to Fourier coefficients $R_k$ using the discrete Fourier transform (DFT).

A synthesized GAW $y_F(\mathrm{t})$ is obtained by using a Fourier synthesizer (FS). The Fourier synthesizer uses the extracted instantaneous phase Θ(t) and the Fourier coefficients $R_k$ to model the input GAW. The synthesized GAW $y_F(\mathrm{t})$ is multiplied with an amplitude modulator $m(t)$ to obtain a modulated GAW $\hat{y}(t).\,m(t)$ is obtained using the minima and maxima extracted from the quasi-unit pulse trains. A GAW obtained using the unmodulated quasi-unit pulse train $u_t$ is the output of a non-modulating model where the frequency and amplitude modulation present in the original GAWs are not modelled. Random modulations of the individual pulses of the input GAW are modelled by modulating the pulse heights and time instants of the quasi-unit pulse trains with reference to the input GAW. The modelled GAW obtained using the modulated quasi-unit pulse trains is the output of the modulating model.

The modelling errors of the two modelled GAWs (modulated and unmodulated) are determined to classify the GAWs into the positive and the negative groups using a simple support vector machine (SVM) classifier. The modelling errors are obtained by taking the root mean squared difference between the input GAW $y(t)$ and the modelled GAWs.

### RESULTS

Two modelled GAWs are obtained for each input GAW. The modelled GAW obtained using the modulating model is observed to fit the input GAW better than the modelled GAW obtained using the non-modulating model. In the non-modulating model, the quasi-unit pulse train fails to track the instantaneous frequency and amplitude modulation of the individual pulses of the input GAW resulting in an estimated modulator with negligible fluctuation. On the other hand, in the modulating model, the modulator estimated using the quasi-unit pulse train fluctuates in accordance with the input GAW which results in a smaller modelling error than the modelling error obtained by using a non-modulating model.

The two modelling errors are the features used for classifying amplitude modulated vocal fry GAWs (positive group) and euphonic (negative group). For euphonic GAWs, the modulation is negliglible as compared to what is seen in vocal fry GAWs. Therefore, the modelling errors for euphonic GAWs obtained using the two models are similar which makes the difference between the modelling errors smaller than the difference between the modelling errors obtained for vocal fry GAWs. As a result, the modelling errors of the euphonic GAWs are well separated from the modelling errors of the vocal fry GAWs in the feature space. The SVM classifier with a linear kernel achieves a perfect 5-fold cross-validated classification accuracy of 100% for both natural and synthetic GAWs. Sensitivities, specificities, and accuracies of classification between vocal fry and euphonic GAWs with 95% confidence intervals (CI) are given in the table below.

**Table 3 Sensitivities, specificities, and accuracies of classification between vocal fry and euphonic GAWs with 95% confidence interval (CI).**

| GAWs | Sensitivity | Specificity | Accuracy |
|---|---|---|---|
| Natural | 59.04% to 100.00% | 63.06% to 100.00% | 78.20% to 100.00% |
| Synthetic | 99.59% to 100.00% | 98.78% to 100.00% | 99.69% to 100.00% |

### DISCUSSION AND CONCLUSION

This paper investigated different types of amplitude modulated vocal fry GAWs. They were modelled using an analysis-by-synthesis approach and distinguished automatically from euphonic GAWs based on their modelling errors. Modulated and unmodulated GAWs were modelled for vocal fry and euphonic GAWs. Modelling errors of natural and synthetic vocal fry GAWs are observed to be well separated from the euphonic GAWs in the feature space. These modelling errors are used as predictors for classifying the vocal fry and euphonic GAWs. For natural and synthetic GAWs, no false positives or false negatives were obtained for classification between vocal fry and euphonic GAWs. Results of classification accuracies suggest that the proposed model enables distinction of normal and vocal fry GAWs. The obtained classification accuracies of detection were found to be competitive with the accuracies reported for past detection techniques proposed in [4, 5 and 6]. Although our model gives a classification with no false positives or false negatives, it only distinguishes between vocal fry and euphonic GAWs. We suggest for the future to add other types of dysphonic voices as negative data. Also, instead of modelling temporal transitions between voice qualities, intervals of homogeneous voice qualities were preselected. Thus, models of temporal transitions of voice qualities related to vocal fry may be proposed in the future.

### ACKNOWLEDGMENTS

This work was supported by the Austrian Science Fund (FWF), KLI722-B30 and the University Hospital Erlangen kindly provided the segmentation tool. Parts of the results were obtained using the Vienna Scientific Cluster (VSC) computational facilities.

### REFERENCES

[1] Whitehead RL, Metz DE, and Whitehead BH. Vibratory patterns of the vocal folds during pulse register phonation. *The Journal of the Acoustical Society of America* 75.4, 1984: 1293-1297.

[2] Blomgren M, Chen Y and Gilbert HR. Acoustic, aerodynamic, physiologic, and perceptual properties of modal and vocal fry registers. *The Journal of the Acoustical Society of America* 103.5, 1998, 2649-2658.

[3] Ishi CT, Ishiguro H and Hagita N. Proposal of acoustic measures for automatic detection of vocal fry. *Ninth European Conference on Speech Communication and Technology*, 2005.

[4] Drugman T, Kane J and Gobl K. Data-driven detection and analysis of the patterns of creaky voice. *Computer Speech & Language*, 2014, 28(5), 1233-1253.

[5] Martin P. Automatic detection of voice creak. *Speech Prosody, Sixth International Conference*, Shanghai, China, 2012**.**

[6] Aichinger P and Pernkopf F. Synthesis and Analysis-by-Synthesis of Modulated Diplophonic Glottal Area Waveforms. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 29 (2021): 914-926.

[7] Devaraj V and Aichinger P. Modelling of Amplitude Modulated Vocal Fry Glottal Area Waveforms Using an Analysis-by-synthesis Approach. *Applied Sciences* 11.5 (2021): 1990.

# MULTI-MODAL EX VIVO SETUP FOR 3D IMAGING OF THE MEDIAL AND SUPERIOR VOCAL FOLD SURFACES IN HEMI LARYNGES

**Reinhard Veltrup[1], Susanne Angerer[1], Marion Semmler[1], Andreas M. Kist[1], Tobias Zillig[1], Julian Zilker[1], Youri Maryn[2], Monique Verguts[2], Joost J.S. van Dinther[2], David. A. Berry[3], Michael Döllinger[1]**

[1] Div. of Phoniatrics and Pediatric Audiology, Dept. of Otorhinolaryngology, University Hospital Erlangen, Germany
[2] European Institute for ORL-HNS, Otorhinolaryngology & Head and Neck Surgery, GZA Sint-Augustinus, Belgium
[3] Department of Head and Neck Surgery, David Geffen School of Medicine at UCLA, Los Angeles, California, USA

***Keywords*: 3D reconstruction; Hemi-larynx; laser projection; high-speed**

## INTRODUCTION

The shape and dynamics of the vocal folds (VF) provide important information for understanding voice production and is part of an ENT physician's examination [1]. For the latter, clinical laryngoscopy is a well-established diagnostic tool. However, it only delivers information about the VF superior surface. Additionally, the relationships between medial and superior VF vibration are largely unknown. In the context of this study, a set-up was developed that allows multi-modal measurements and analysis of both the superior and medial VF surface. It enables us to record high-speed videos of the *ex vivo* larynx from superior and medial views and synchronously measure the subglottal pressure and the emitted acoustic signal. From the video data sets, the superior and medial VF surfaces are reconstructed and afterwards merged into one complete VF 3D-surface. In total, ten human larynges were systematically investigated and evaluated for six different pre-phonatory settings.

## METHODS

The 3D printed test stand, which holds the hemi-larynx, allows physiological movement of the thyroid cartilage. With sutures and weights, the adduction and elongation processes are simulated. The video data was recorded with two high-speed cameras at 4000 frames per second. From the camera data, the superior and medial VF surfaces are reconstructed separately in two different ways: (1) superior surface with a laser projection unit according to Semmler *et al* [2], (2) medial surface by stereo vision with a prism and sewn-in marker points according to Döllinger *et al* [1]. The detection of both the laser and marker points are performed with artificial neural network based in-house tools. Subsequently, the two partial surfaces are merged into one complete VF surface. Each larynx was measured with two elongation steps and three adduction increments (elongation: 10 g, 20 g; adduction: 10 g, 20 g, 50 g). At each of these settings, five measurements were performed at equidistant airflow rates, starting from phonation onset.

## RESULTS

The camera data shows that the VFs perform physiological mucosal wave motions. Oscillation frequencies, mean subglottal pressure and threshold onset pressure are within physiological ranges [3]. At a defined pre-phonatory setting, the linear increase of the airflow rate causes a linear increase in oscillation frequency and the mean subglottal pressure. With minimal elongation and adduction levels, the mean fundamental frequency between minimum and maximum flow rates increases from 112 Hz to 139 Hz for male larynges and 248 Hz to 294 Hz for female larynges. Medial and superior surfaces were reconstructed successfully (Figure 1). Integrating both data sets led to a complete dynamic VF surface for a sequence of 10 consecutive oscillation periods.

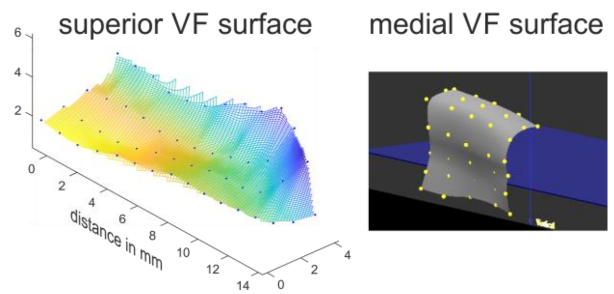



Figure 1. reconstructed superior und medial VF surfaces.

## CONCLUSION

The developed test stand allows for a multi-modal analysis of hemi-larynges with respect to 3D dynamics of the VF. The next aim is to investigate the relationship between the dynamics of the medial and superior VF surface in interaction with the acoustic signal and the subglottal pressure.

## REFERENCES

[1] Döllinger *et al*, J Voice, 20:401-413, 2006
[2] Semmler *et al*, IEEE-TMI, 35:1615-1624, 2016
[3] I. R. Titze, Principles of Voice Production, XIX, 2000

# PHYSICS OF PHONATION OFFSET: TOWARDS UNDERSTANDING RELATIVE FUNDAMENTAL FREQUENCY OBSERVATIONS

**Mohamed Serry[1], Cara E. Stepp[2], & Sean D. Peterson[1]**

[1]Mechanical and Mechatronics Engineering Department, University of Waterloo, Waterloo, ON, N2L 3G1, Canada
[2]Department of Speech, Language & Hearing Sciences, Boston University, Boston, MA 02215, USA

***Keywords*: Relative Fundamental Frequency; Vocal Hyperfunction; Phonation Offset; Reduced-Order Models**

## INTRODUCTION

Relative fundamental frequency (RFF) characteristics during phonation periods surrounding voiceless consonants differ between normal and hyperfunctional speakers, making RFF a viable classification tool [1]. However, RFF measurements are generally prone to several sources of inter- and intra-subject variability, which limits its assessment capability. Comprehensive understanding of the underlying physics of RFF can potentially elucidate some of the underlying mechanisms of vocal hyperfunction and make RFF assessment capabilities more robust. The objective of this work is to analyze the underlying mechanisms associated with the observed reduction in RFF during phonation offset in both normal and hyperfunctional speakers.

## METHODS

First, we introduce a quasi-steady impact oscillator model that abstracts the mechanics of vocal fold (VF) vibrations during offset and enables analytical treatment. The theoretical study is followed by extensive numeral simulations, wherein a body-cover phonation model [2] incorporating muscle activation rules [3] and a time-varying glottal gap, tuned based on glottal angle empirical data [4] to capture VF abduction, is employed.

## RESULTS

Theoretical analysis shows that fundamental frequency (FF) is influenced by degree of VF collision; that is, as the glottal gap increases there is a decrease in collision, and FF decreases, in agreement with clinical RFF observations. Numerical simulations verify the decrease in FF correlates with a drop in collision forces ($F_{col-}$) during offset, as seen in Figure 1. At the cessation of collision, the system stiffness no longer contains a contribution from collision and FF increases to the damped natural frequency of the system. Numerical simulations reveal the sensitivity of RFF patterns to the abduction initiation time relative to the phonation cycle, and the length of abduction period, which may contribute to the variability of clinical RFF measurements. Furthermore, the potential role of the cricothyroid muscle in stabilizing FF during offset is identified, with relative activation levels of the muscle eliciting different RFF patterns. Specifically, coordinated increase in cricothyroid activation during abduction mitigates the drop in RFF during offset. Conversely, starting with stronger lateral cricothyroid activation at abduction initiation increases the RFF drop.

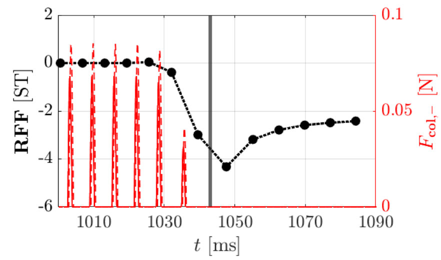



Figure 1. RFF and collision forces over time. The thick vertical line indicates collision cessation.

## CONCLUSION

The drop in RFF during offset can be attributed to the decrease in collision forces associated with the abduction process, where RFF patterns exhibit sensitivity towards the temporal characteristics of abduction. In general, the degree of drop depends on the pre-offset collision levels and the activation levels of the *cricothyroid* muscle during abduction, which can explain, in part, the differences in RFF patterns between healthy and hyperfunctional speakers.

## ACKNOWLEDGMENTS

Research reported in this work was supported by the NIDCD of the NIH under awards P50DC015446 and R01DC015570. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

## REFERENCES

[1] Stepp, C. E., Hillman, R. E., & Heaton, J. T. (2010). The Impact of Vocal Hyperfunction on Relative Fundamental Frequency During Voicing Offset and Onset. *J Speech Lang Hear Res*, **53**, 1220-1226.

[2] Story, B. H., & Titze, I. R. (1995). Voice simulation with a body-cover model of the vocal folds. *J Acoust Soc Am*, **97**(2), 1249-1260.

[3] Titze, I. R., & Story, B. H. (2002). Rules for controlling low-dimensional vocal fold models with muscle activation. *J Acoust Soc Am*, **112**(3), 1064-1076.

[4] Diaz-Cadiz, M., McKenna, V. S., Vojtech, J. M., & Stepp, C. E. (2019). Adductory Vocal Fold Kinematic Trajectories During Conventional Versus High-Speed Videoendoscopy. *J Speech Lang Hear Res*, **62**(6), 1685-1706.

# ALTERNATIVE VOICE THERAPY METHODS FOR CLINICAL PRACTICE

**Edwin Yiu (PhD)**

Professor of Speech Pathology, The University of Hong Kong

Vocal hygiene and vocal exercises are traditional voice treatment methods. Many of these methods are based on classical vocal pedagogies corroborated with modern scientific validations (e.g., resonant voice therapy, vocal function exercise). These orthodox voice treatment methods dominate the literature as well as in clinical practice. Researchers, however, are constantly exploring for new treatment methods for voice disorders. This talk will introduce two alternative voice treatment approaches: acupuncture and vibrational therapy (localized vibration and whole-body vibration). These have been found to be effective in treating vocal fold lesions related with phonotrauma and vocal fatigue. The principles, procedures, and the validation processes of these two treatment approaches will be presented in this talk.

# RELATIONSHIPS BETWEEN DAILY SPEAKING VOICE USE AND PERSONALITY IN SINGERS WITH HEALTHY VOICES

**Laura E. Toles[1,2], Nelson Roy[4], Stephanie Sogg[2,3], Katherine L. Marks[1,2], Andrew J. Ortiz[2], Daryush D. Mehta[1,2,3], Robert E. Hillman[1,2,3]**

[1] MGH Institute of Health Professions, Charlestown, MA, USA
[2] Massachusetts General Hospital, Boston, MA, USA
[3] Harvard Medical School, Boston, MA, USA
[4] University of Utah, Salt Lake City, UT, USA

***Keywords*: Voice; Ambulatory Voice Monitoring; Personality; Singers**

## INTRODUCTION

Singers are at an elevated risk of developing phonotrauma [1]. Recent work has found that singers with phonotrauma speak more than vocally healthy singers, but do not sing more [2]. The authors hypothesized that increased speaking voice use, combined with the tendency to speak with higher laryngeal forces and more abrupt/pressed glottal closure (suggesting more frequent use of louder voice), might represent etiological factors that are influenced by the intrinsic personality dispositions of singers. This study sought to determine whether personality traits of vocally healthy singers are related to their speaking behaviors in daily life that could be associated with increased risk for vocal fold trauma (e.g., speaking louder, higher vocal doses).

## METHODS

Weeklong ambulatory voice recordings using a neck-placed accelerometer were obtained for 47 vocally healthy female singers. A singing classifier [3] was applied to the acceleration signal, and singing was removed from the analysis to address our hypothesis that speaking behaviors are more likely to be associated with personality. Each participant also completed the Multidimensional Personality Questionnaire – brief form, which measures eleven trait facets. Relationships between personality facets and speaking voice (vocal dose, sound pressure level (SPL), fundamental frequency ($f_o$)) distributional parameters (mean, standard deviation, range, skewness, kurtosis) were examined using pairwise Pearson $r$ correlation coefficients. Subsequent multiple regression was performed for voice parameters that correlated significantly with two or more trait facets ($p < .05$).

## RESULTS

Each multiple regression model was statistically significant, with the overall multiple $R$ ranging from .39 to .54. Vocal dose measures were positively correlated with a combination of Wellbeing and Social Potency facets and negatively correlated with Absorption. Mean SPL was positively correlated with Wellbeing and negatively correlated with Absorption. SPL variability and range measures were positively correlated with Social Potency and negatively correlated with Harm Avoidance. None of the personality traits were significantly correlated with $f_o$ measures.

## DISCUSSION & CONCLUSION

Personality traits are related to daily speaking voice use in vocally healthy singers. Individuals with higher levels of traits related to happiness and social dominance and lower propensities to use caution tended to speak more, more loudly, and with more SPL variability, which could theoretically increase risk of phonotrauma. Future work will investigate these relationships in singers with phonotrauma.

## ACKNOWLEDGMENTS

Funding provided by the National Institutes of Health (NIH) National Institute on Deafness and other Communication Disorders under grants P50 DC015446 (PI: Hillman) and F31 DC019284-01 (PI: Toles). The paper's contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH. Special thanks to Dr. Annie B. Fox for her assistance with statistical analysis.

## REFERENCES

[1] Verdolini & Ramig. Review: Occupational risks for voice problems. *Logopedics, Phoniatrics, Vocology*, 2001, 26, 37-46.

[2] Toles et al. Differences between female singers with phonotrauma and vocally healthy matched controls in singing and speaking voice use during 1 week of ambulatory monitoring. *American Journal of Speech-Language Pathology*, 2021, 1-11.

[3] Ortiz et al. Automatic speech and singing classification in ambulatory recordings for normal and disordered voices. *Journal of the Acoustical Society of America*, 2019, 146, EL22-EL27.

# THE RELATIONSHIP BETWEEN VOICE ONSET TIME AND INCREASES IN VOCAL EFFORT AND FUNDAMENTAL FREQUENCY

**Matti D. Groll[1,2], Surbhi Hablani[2], Cara E. Stepp[1,2,3]**

[1] Department of Biomedical Engineering, Boston University, Boston, MA, USA
[2] Department of Speech, Language, and Hearing Sciences, Boston University, Boston, MA, USA
[3] Department of Otolaryngology – Head and Neck Surgery, Boston University School of Medicine, Boston, MA, USA

***Keywords*: Voice Onset Time; Vocal Hyperfunction; Vocal Effort; Vocal Strain**

## INTRODUCTION

Prior work suggests that voice onset time (VOT) may be impacted by laryngeal tension: VOT means decrease when individuals with typical voices increase their fundamental frequency ($f_o$) [1] and VOT variability is increased in individuals with vocal hyperfunction [2], a voice disorder characterized by increased laryngeal tension. This study further explored the relationship between VOT and laryngeal tension during increased $f_o$, vocal effort, and vocal strain.

## METHODS

Sixteen typical speakers of American English were instructed to produce VOT utterances under four conditions: baseline, high pitch, effort, and strain. Various VOT utterances were produced in the carrier sentence "Say /vowel-consonant-vowel/ again" for /a/ and /u/ vowels using each of the six American English stop consonants. Each unique VOT utterance was repeated three times, resulting in a set of 36 utterances (6 consonants x 2 vowels x 3 repetitions). Repeated measures analysis of variance models was used to analyze the effects of voicing, place of articulation, vowel, and condition on VOT means and standard deviations (SDs); pairwise comparisons were used to determine significant differences between conditions.

## RESULTS

Voicing, condition, and their interaction significantly affected VOT means. Voiceless VOT means significantly decreased for high pitch ($p < .001$) relative to baseline; however, no changes in voiceless VOT means were found for effort or strain relative to baseline. Although condition had a significant effect on VOT SDs, there were no significant differences between effort, strain, and high pitch conditions relative to baseline.

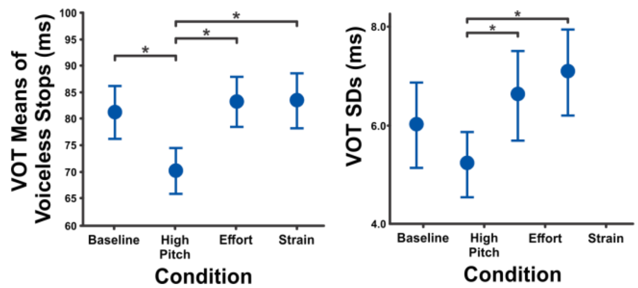



Figure 2. Average voice onset time (VOT) means and standard deviations (SDs) and 95% confidence intervals for each experimental condition. Brackets indicate significant differences between conditions ($p < .05$).

## DISCUSSION AND CONCLUSION

Speakers with typical voices likely engage different musculature to increase pitch than to increase vocal effort and strain. The increased VOT variability present with vocal hyperfunction is not seen in individuals with typical voices using increased effort and strain, supporting the assertion that this feature of vocal hyperfunction may be related to disordered vocal motor control rather than resulting from effortful voice production.

## ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health grants R01 DC015570 (CES), P50 DC015446 (REH), and T32 DC013017 (CAM) from the National Institute on Deafness and Other Communication Disorders (NIDCD).

## REFERENCES

[1] McCrea, C. R., & Morris, R. J. (2005). The effects of fundamental frequency level on voice onset time in normal adult male speakers. *Journal of Speech, Language, and Hearing Research, 48*(5), 1013-1024.

[2] McKenna, V. S., Hylkema, J. A., Tardiff, M. C., & Stepp, C. E. (2020). Voice onset time in individuals with hyperfunctional voice disorders: Evidence for disordered vocal motor control. *Journal of Speech, Language, and Hearing Research, 63*(2), 405-420.

# LOMBARD EFFECT RETENTION AFTER NOISE REMOVAL IN PATIENTS WITH PHONOTRAUMATIC AND NON-PHONOTRAUMATIC VOCAL HYPERFUNCTION

**C. Castro[1,2,3], P. Prado[4], V. M. Espinoza[5], J. P. Cortés[6], R. Manríquez[1], A. Testart[7], M. Zañartu[1]**

[1] Department of Electronic Engineering, Universidad Técnica Federico Santa María, Valparaíso, Chile
[2] Department of Communication Disorders, Universidad de Chile, Santiago, Chile
[3] Department of Communication Disorders, Universidad de Valparaíso, Valparaíso, Chile
[4] Latin American Brain Health Institute (BrainLat), Universidad Adolfo Ibáñez, Santiago de Chile, Chile
[5] Department of Sound, Universidad de Chile, Santiago, Chile
[6] Center for Laryngeal Surgery and Voice Rehabilitation Laboratory, Massachusetts General Hospital, Boston, Massachusetts
[7] Department of Communication Disorders, Universidad de Playa Ancha, Valparaíso, Chile

***Keywords*: Lombard Effect, Acoustic measure, Accelerometer, Auditory feedback, Voice production**

## INTRODUCTION

Vocal Hyperfunction (VH) refers to excessive perilaryngeal musculoskeletal activity [1]. A recent framework proposes two subtypes of VH depending on the presence or absence of vocal fold tissue lesion, i.e., phonotraumatic VH (PVH) and non-phonotraumatic VH (NPVH) [1]. Previous studies using auditory feedback perturbation with pitch shift paradigm have shown that relevant features of auditory-motor function are impaired in some individuals with VH [2],[3]. We propose extending these efforts by evaluating the vocal function when the SPL of the vocal target is involuntarily modulated by the background noise.

## METHODS

*Participants*

10 participants with PVH, 10 participants with NPVH, and 10 participants with typical voices were recruited for this study. All participants are not singers, presented normal threshold assessment by clinical pure-tone audiometry, and did not have a history of speech disorders.

*Instrumentation and Measurement*

All subjects were asked to utter a series of 5 vowels and 10 syllables /pae/, each lasting 3 seconds with a 3-second pause between them, under three conditions, baseline (in quiet), Lombard (speaking while listening to noise), and recovery (5 minutes after the end of LE conditions). The Lombard condition included an induction stage; the participants were instructed to produce fifty vowels /æ/ or /i/ in a pseudorandom order to attune the change in the acoustic environment. The LE was elicited with speech noise presented at 80 dB. The vocal function was assessed by acoustic, aerodynamic, and accelerometer measurement.

The acoustic signal was obtained using a condenser microphone (4961, B&K). Aerodynamic signals were recorded using a circumferentially vented mask (MS-110, Glottal Inc,) and a neck-surface accelerometer (BU-2713, Knowles).

*Analysis*

Mean voice intensity (SPL), the difference between first and second harmonic amplitudes (H1-H2), peak-to-peak amplitude of the unsteady airflow (AC Flow), maximum flow declination rate (MFDR), and open quotient (OQ) was estimated using custom MATLAB scripts. Two-way mixed ANOVAs ($p<0.05$) were performed to analyze the dynamics of the different parameters, using the group and the experimental condition as factors.

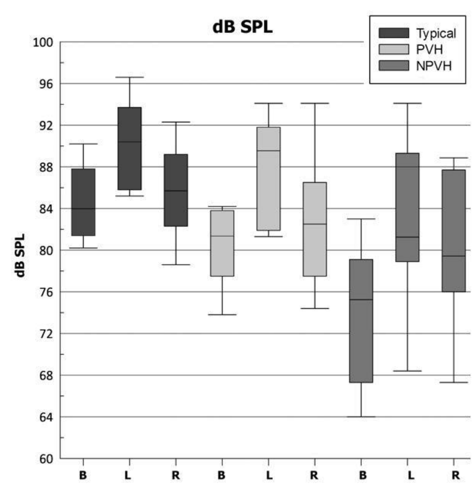



Figure 1. Mean voice intensity (SPL) for Typical, PVH and NPVH group in Baseline, Lombard, and Recovery conditions.

## RESULTS

All three groups exhibited a significant increase in SPL, MFDR, ACFlow, and SGP as well as a decrease in OQ and H1-H2 when speaking in noise (Lombard effect). However, only individuals with healthy and PVH voices returned to baseline conditions for the Lombard effect removal, i.e., when speaking in quiet after five minutes of recovery. Individuals NPVH exhibited significant differences between baseline and recovery conditions for SPL, MFDR, ACFlow, and OQ.

## DISCUSSION

Preliminary results suggest that subjects with NPVH have more difficulties returning to baseline conditions once the noise is removed in comparison with participants with typical voices and participants with PVH. These results are consistent with those reported in previous studies using the pitch shift paradigm on VPH [2]. Based on proposed by internal models for speech motor control [4], [5],[6]. The persistence of the Lombard effect after the 5 minutes of removing the noise could be related to difficulties on reprogramming of feedforward commands in individuals with NPVH when speaking in noise.

.

## CONCLUSION

Individuals with NPVH in this study exhibited an after-effect of speaking in noise with a retain the Lombard effect even after the removal of noise, this could be related with difficulties updating the feedforward commands.

## 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] Hillman, R. E., Stepp, C. E., Van Stan J. H., Zañartu, M., & Mehta, D. D. (2020). An Updated Theoretical Framework for Vocal Hyperfunction. American Journal of Speech-Language Pathology, 29(4), 2254-2260

[2] Stepp, C. E., Lester-Smith, R. A., Abur, D., Daliri, A., Pieter N., J., & Lupiani, A. A. (2017). Evidence for Auditory-Motor Impairment in Individuals With Hyperfunctional Voice Disorders. Journal of Speech, Language, and Hearing Research: JSLHR, 60(6), 1545–1550.

[3] Lee, S. H., Yu, J. F., Fang, T. J., & Lee, G. S. (2019). Vocal fold nodules: A disorder of phonation organs or auditory feedback? Clinical otolaryngology: official journal of ENT-UK; official journal of Netherlands Society for Oto-Rhino-Laryngology & Cervico-Facial Surgery, 44(6), 975–982.

[4] Tourville, J. A., & Guenther, F. H. (2011). The DIVA model: A neural theory of speech acquisition and production. Language and Cognitive Processes, 26(7), 952-981.

[5] Parrell, B., Ramanarayanan, V., Nagarajan, S., & Houde, J. (2019). The FACTS model of speech motor control: Fusing state estimation and task-based control. PLoS computational biology, 15(9), e1007321.

[6] Houde, J. F., & Nagarajan, N. S. (2011). Speech production as state feedback control. Frontiers in Human Neuroscience, 5, 82.

# ANALYSIS AND COMPARISON OF DIFFERENT BIOMECHANICAL MODELS OF VOCAL PRODUCTION

**Carlos Calvache[1, 2], Leonardo Solaque[1], Alexandra Velasco, Lina Peñuela[1]**

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

***Keywords*: Voice; Vocal Production models; Modelling; Vocal Folds**

## INTRODUCTION

The acoustic signal of the human voice is the result of a combination of three main components: (1) aerodynamic variables such as air flow and pressure (subglottic level); (2) vibration of the vocal folds (glottal level); and (3) sound propagated in the vocal tract (supraglottic level) [1].

From a clinical voice perspective, it is necessary to understand the vocal production physiology. Mathematical models allow identifying the aerodynamic, mechanical, and acoustic variables involved [2]. The use of mathematical models can help clinicians to determine parameters that are not easily assessed due to the biomechanical nature of the larynx, especially the parameters that cannot be directly measured, such as impact or collision stress of vocal folds during its vibration [1], [3]. This feature is of great relevance for the diagnosis and rehabilitation of hyperfunctional vocal pathologies. Despite the evidence, there are still many challenges in identifying and measuring several characteristics of the physical phenomena involved in vocal physiology [4], [5]. Thus, this study aims to determine the physiological characteristics to define a biomechanical model that allows understanding mechanical stress on the vocal cords.

## METHODS

A descriptive-deductive study is used to analyze and compare different biomechanical vocal production models. The content analysis technique is the main tool of this research. The analysis was based on a systematic review of the literature carried out by the authors of the present work [5]. The content analysis is focused on defining the physiological characteristics of a biomechanical model that allows quantifying the impact forces on the vocal folds during their vibration.

### *Relational Analysis*

The following steps were developed:

1. Correlational analysis between the different types of models found in the literature (Taxonomy) [5] and physiological characteristics of vocal production.
2. Proximity analysis through the evaluation of concepts co-occurrence. The co-occurrence network reported in systematic review of the literature [5] was taken into account.
3. Statement of the relational analysis question based on the proximity analysis.
4. Definition of the classifiers and the analysis categories to reach conclusions with the analyzed literature.

## RESULTS AND DISCUSSION

The models found in literature were classified into the following categories: (1) Models that represent the source (vocal cords); (2) Models representing the filter (vocal tract); (3) Models that represent the source-filter interaction; and (4) models that represent the airflow-source interaction.

The proximity analysis shows models that represent the source have correlations between phonation and different biomechanical processes of the vocal folds; models that represent filters identify processes associated with resonance and voice like acoustic phenomena. Models of source-filter represent the flow-structure interaction and related features to the deformation of the vocal folds, their displacement, and the acceleration concerning aerodynamic forces. The proximity analysis also shows concepts associated with the acoustic qualities of voice.

Based on this semantic analysis, we state the relational analysis question: *Which are the physiological characteristics and the physical phenomena referred to in the analyzed models?* This question served as the basis for determining the final analysis categories. Thus, we found that a biomechanical model of vocal production, where the magnitude of the impact forces of the vocal folds identified, must consider:

1. The nonlinear anisotropic viscoelastic behavior of the vocal folds[6], [7].
2. The representation of a self-oscillating acoustic aerodynamic conversion system [8], [9].
3. The relationship between the geometry of the vocal folds, the transglottal flow, and the inertia of vocal tract [10], [11].

4. The arytenoid adduction shape is important to define complete glottic closure [2], [12], [13].
5. Subglottic pressure magnitude is directly related to the intraglottal forces, and therefore, to the temporal pattern of the waveform of the glottic area [8], [14], [15].
6. Impact forces between vocal folds depend on the magnitude of transglottal flow and the damping / stiffness constants of vocal folds. [16], [17] .

## CONCLUSION

The evidence allows understanding that vocal folds' mechanical conditions are affected at the contact surface during their self-oscillation (Mucosa). Intraglottal impact forces become the main cause of phono-traumatic injuries. The most relevant models for the vocal clinic explain how the geometry of the vocal cords is affected by the flow behavior, which is a key point in the physiological characteristics to define a biomechanical model of voice production.

## REFERENCES


[1] Z. Zhang, "Mechanics of human voice production and control," *J. Acoust. Soc. Am.*, vol. 140, no. 4, pp. 2614–2635, 2016, doi: 10.1121/1.4964509.

[2] R. Schwarz *et al.*, "Modeling the pathophysiology of phonotraumatic vocal hyperfunction with a triangular glottal model of the vocal folds," *J. Acoust. Soc. Am.*, vol. 139, no. 2, pp. 1–16, Aug. 2016, doi: 10.1121/1.1577547.

[3] D. D. Mehta, M. Zañartu, S. W. Feng, H. A. I. Cheyne, and R. E. Hillman, "Mobile voice health monitoring using a wearable accelerometer sensor and a smartphone platform," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 12 PART2, pp. 3090–3096, 2012, doi: 10.1109/TBME.2012.2207896.

[4] M. Döllinger, S. Kniesburges, M. Kaltenbacher, and M. Echternach, "Aktuelle Methoden zur Modellierung des Stimmgebungsprozesses," *HNO*, vol. 64, no. 2, pp. 82–90, 2016, doi: 10.1007/s00106-015-0110-x.

[5] C. Calvache, L. Solaque, A. Velasco, and L. Peñuela, "Biomechanical Models to Represent Vocal Physiology: A Systematic Review," *J. Voice*, 2021, doi: 10.1016/j.jvoice.2021.02.014.

[6] F. Alipour, D. A. Berry, and I. R. Titze, "A finite-element model of vocal-fold vibration," *J. Acoust. Soc. Am.*, vol. 108, no. 6, pp. 3003–3012, Dec. 2000, doi: 10.1121/1.1324678.

[7] Z. Zhang, "Structural constitutive modeling of the anisotropic mechanical properties of human vocal fold lamina propria," *J. Acoust. Soc. Am.*, vol. 145, no. 6, pp. EL476–EL482, Jun. 2019, doi: 10.1121/1.5109794.

[8] J. Horáček, A. M. Laukkanen, and P. Šidlof, "Estimation of impact stress using an aeroelastic model of voice production," *Logop. Phoniatr. Vocology*, vol. 32, no. 4, pp. 185–192, 2007, doi: 10.1080/14015430600628039.

[9] I. R. Titze, "Where has all the power gone? Energy production and loss in vocalization," *SPEECH Commun.*, vol. 101, pp. 26–33, Jul. 2018, doi: 10.1016/j.specom.2018.05.003.

[10] I. R. Titze, "Theory of glottal airflow and source-filter interaction in speaking and singing," *ACTA Acust. UNITED WITH Acust.*, vol. 90, no. 4, pp. 641–648, 2004.

[11] B. H. Story, "An overview of the physiology, physics and modeling of the sound source for vowels," *Acoust. Sci. Technol.*, vol. 23, no. 4, pp. 195–206, 2002, doi: 10.1250/ast.23.195.

[12] G. E. Galindo, S. D. Peterson, B. D. Erath, C. Castro, R. E. Hillman, and M. Zañartu, "Modeling the pathophysiology of phonotraumatic vocal hyperfunction with a triangular glottal model of the vocal folds," *J. Speech, Lang. Hear. Res.*, vol. 60, no. 9, pp. 2452–2471, 2017, doi: 10.1044/2017_JSLHR-S-16-0412.

[13] M. Zanartu, G. E. Galindo, B. D. Erath, S. D. Peterson, G. R. Wodicka, and R. E. Hillman, "Modeling the effects of a posterior glottal opening on vocal fold dynamics with implications for vocal hyperfunction," *J. Acoust. Soc. Am.*, vol. 136, no. 6, pp. 3262–3271, 2014, doi: 10.1121/1.4901714.

[14] J. Horáček, P. Šidlof, and J. G. Švec, "Numerical simulation of self-oscillations of human vocal folds with Hertz model of impact forces," *J. Fluids Struct.*, vol. 20, no. 6 SPEC. ISS., pp. 853–869, 2005, doi: 10.1016/j.jfluidstructs.2005.05.003.

[15] B. D. Erath, S. D. Peterson, K. S. Weiland, M. W. Plesniak, and M. Zañartu, "An acoustic source model for asymmetric intraglottal flow with application to reduced-order models of the vocal folds," *PLoS One*, vol. 14, no. 7, pp. 1–27, 2019, doi: 10.1371/journal.pone.0219914.

[16] W. Jiang, X. Zheng, and Q. Xue, "Influence of vocal fold cover layer thickness on its vibratory dynamics during voice production," *J. Acoust. Soc. Am.*, vol. 146, no. 1, pp. 369–380, Jul. 2019, doi: 10.1121/1.5116567.

[17] J. Horacek, A.-M. Laukkanen, P. Sidlof, P. Murphy, and J. G. Svec, "Comparison of Acceleration and Impact Stress as Possible Loading Factors in Phonation: A Computer Modeling Study," *FOLIA Phoniatr. Logop.*, vol. 61, no. 3, pp. 137–145, 2009, doi: 10.1159/000219949.

# CLASSIFICATION OF VOCAL FATIGUE USING NECK SEMG WITH LEAVE-ONE-SUBJECT-OUT TESTING

**Yixiang Gao[1], Maria Dietrich[2,3], G. N. DeSouza[1]**

[1] Department of Electrical Engineering and Computer Science, University of Missouri, Columbia, MO, United States
[2] Department of Psychiatry and Psychotherapy, University Hospital Bonn, Bonn, Germany
[3] Department of Speech, Language and Hearing Sciences, University of Missouri, Columbia, MO, United States

***Keywords*: Voice; Surface Electromyography; Pattern Recognition; Vocal Fatigue**

## INTRODUCTION

Vocal fatigue is a leading vocal symptom among teachers that can threaten a teacher's career [1]. Our previous study using surface electromyography (sEMG) to classify vocal fatigue has shown promising results to adapt such techniques in a clinical setting [2]. This work presents a detailed investigation using **leave-one-out (LOO)** experiments to highlight challenges of this machine learning approach.

## METHODS

*Participants*

We age matched a balanced group of 40 subjects, 20 vocally healthy (non-teachers) and 20 vocally fatigued (early career teachers), from a total of 88 subjects. The descriptive statistics for both groups are shown in Table **4**. All vocally healthy and vocally fatigued subjects had a VFI-1 $\leqslant$ 10 and VFI-1 $\geqslant$ 15 respectively [3].

**Table 4. Descriptive statistics for 40 test subjects matched on age and neck skinfold thickness.**

| | Vocally Fatigued | Vocally Healthy |
|---|---|---|
| # of subjects | 20 | 20 |
| Age (years) | 25.6 ± 4.3 | 25.3 ± 4.7 |
| Suprahyoid (mm) | 7.0 ± 3.4 | 5.4 ± 1.3 |
| Infrahyoid (mm) | 6.4 ± 3.1 | 5.1 ± 1.3 |
| VFI-1 | 18.2 ± 5.4 | 2.1 ± 1.7 |

*Instrumentation and Measurement*

For the sEMG signal acquisition of this study, we used a base station and four wireless TrignoTM mini sEMG sensors with a bandwidth of 20 Hz to 450 Hz (Delsys, Natick, MA). More details can be found in [2].

*Analysis*

We conducted our analysis on three vowels at normal pitch and loudness (/a/, /u/, and /i/). We labeled the vowels produced by vocally fatigued/healthy subjects as positive/negative samples, respectively. The total number of positive and negative samples were 3270 and 3202. We used a leave-one-out (LOO) approach, where if $N$ was the number of vocally healthy subjects and $M$ the number of vocally fatigued subjects, a total of $M + N - 1$ subjects were successively selected for training and validation, while the subject left-out was used for testing.

## RESULTS AND DISCUSSION

Our classification can achieve high validation accuracy (97.51%) but the averaged testing accuracy for left-out subjects is 62.36%. However, if we randomly select 10% of the samples (i.e., between 10 and 15 samples) from the testing subject to be included in training and validation, the testing accuracy improves from 62.36% to 83.93%.

**Table 2. Confusion matrix for the positive vs. negative detection of vocal fatigue among 40 matched subjects using leave-one-out approach.**

| | Actual Positive | Actual Negative |
|---|---|---|
| Predicted Positive | 97.52% | 2.51% |
| Predicted Negative | 2.48% | 97.49% |
| Validation Accuracy | 97.51% | |

## CONCLUSION

Classification using LOO has proven to be very challenging. Most of the work presented to date in the literature has failed to cover LOO testing. These two observations raise an important question regarding what the systems are really learning: patterns of vocal fatigue/healthy voice or patterns of subjects in sEMG signals. Therefore, further investigation is required, probably on a larger data set.

## ACKNOWLEDGMENTS

Research funded by NIDCD research grants R15DC015335 and R01DC018026. Special thanks to Ashton Bernskoetter, Taylor Hall, Katherine Johnson, Haley McCabe, Melinda Pfeiffer, and Allison Walker for help with data collection and Matthew Page for reviewing laryngeal videostroboscopies.

## REFERENCES

[1] Smith E, Verdolini K, Gray S, Nichols S, Lemke J, Barkmeier J, Dove H, Hoffman. H. Effect of voice disorders on quality of life. *J Med Speech-Lang Pathol* 1996; 4:223–244.

[2] Gao Y, Dietrich M, DeSouza GN. Classification of vocal fatigue using sEMG: Data imbalance, normalization, and the role of Vocal Fatigue Index scores. *Applied Sciences* 2021; 11(10):4335.

[3] Nanjundeswaran C, Jacobson BH, Gartner-Schmidt J, Verdolini Abbott K. Vocal Fatigue Index (VFI): Development and validation. *J Voice* 2015; 29:433–440.

*The 14th International Conference on Advances in Quantitative Laryngology, Voice and Speech Research (AQL)*

# EXPLORE VOICE PRODUCTION VARIABILITY THROUGH NECK SEMG CLUSTERING - CHALLENGE FOR ACCURATE LABELING OF VOCAL FATIGUE

**Yixiang Gao[1], Maria Dietrich[2,3], G. N. DeSouza[1]**

[1] Department of Electrical Engineering and Computer Science, University of Missouri, Columbia, MO, United States
[2] Department of Psychiatry and Psychotherapy, University Hospital Bonn, Bonn, Germany
[3] Department of Speech, Language and Hearing Sciences, University of Missouri, Columbia, MO, United States

***Keywords*: Voice; Surface Electromyography; Unsupervised Learning; Clustering**

## INTRODUCTION

Supervised learning algorithms may be compromised by assuming that all samples, in repeated trials performed by each subject follows the same class label as the subjects themselves. In practice, not all samples from a subject will present with either the same patterns or the same intensity in the patterns as a simple, crisp label would imply. In this study, we applied different unsupervised learning methods to investigate, based on per-sample-level, the variations within each subject as well as how they correlate with vocal fatigue.

## METHODS

### *Participants*

We age matched a balanced group of 40 subjects, 20 vocally healthy (non-teachers) and 20 vocally fatigued (early career teachers), from a total of 88 subjects. The descriptive statistics for both groups are shown in Table **4**. All vocally healthy and vocally fatigued subjects had a Vocal Fatigue Index (VFI) score on the first factor tiredness of voice of VFI-1 ⩽ 10 and VFI-1 ⩾ 15 [3].

**Table 5. Descriptive statistics for 40 matched test subjects.**

| | Vocally Fatigued | Vocally Healthy |
|---|---|---|
| # of subjects | 20 | 20 |
| Age (years) | 25.6 ± 4.3 | 25.3 ± 4.7 |
| Suprahyoid (mm) | 7.0 ± 3.4 | 5.4 ± 1.3 |
| Infrahyoid (mm) | 6.4 ± 3.1 | 5.1 ± 1.3 |
| VFI-1 | 18.2 ± 5.4 | 2.1 ± 1.7 |

### *Instrumentation and Measurement*

For the sEMG signal acquisition of this study, we used a base station and four wireless TrignoTM mini sEMG sensors with a bandwidth of 20 Hz to 450 Hz (Delsys, Natick, MA). More details can be found in [2].

### *Analysis*

We conducted our analysis on three vowels at normal pitch and loudness (/a/, /u/, and /i/). The unsupervised learning techniques we used did not require any label from the samples. Thus, a total of 6472 samples were fed into the algorithm. We applied both K-Means and Hierarchical Clustering methods on either selected or entire feature set extracted from the samples. Then we analyzed the cluster composition by their vowel classes, average VFI-1 values as well as belonging subjects.

## RESULTS

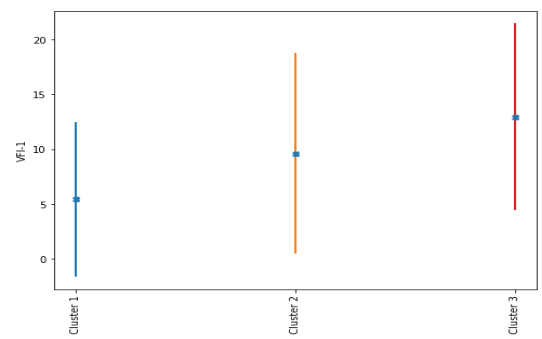



Figure 1. Three clusters returned by K-means and their mean(std) of VFI-1 scores.

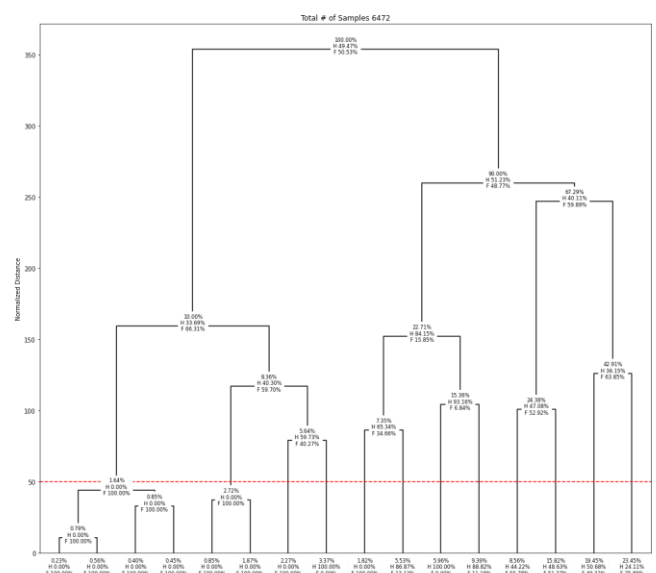



Figure 2. Dendrogram plotted with hierarchical clustering which demonstrated a tree of cluster hierarchy at different thresholds.

### DISCUSSION AND CONCLUSION

The proposed clustering approaches revealed a trend on individual sEMG signals that does not necessarily comply with self-assigned labels (VFI-1). Indeed, clinical reasoning cannot dispute the findings by our unsupervised method that sample labeling should not be extrapolated from patient "macro" conditions.

### ACKNOWLEDGMENTS

Research funded by NIDCD research grants R15DC015335 and R01DC018026. Special thanks to Ashton Bernskoetter, Taylor Hall, Katherine Johnson, Haley McCabe, Melinda Pfeiffer, and Allison Walker for help with data collection and Matthew Page for reviewing laryngeal videostroboscopies.

### REFERENCES

[1] Smith E, Verdolini K, Gray S, Nichols S, Lemke J, Barkmeier J, Dove H, Hoffman. H. Effect of voice disorders on quality of life. *J Med Speech-Lang Pathol* 1996; 4:223–244.

[2] Gao Y, Dietrich M, DeSouza GN. Classification of vocal fatigue ssing sEMG: Data imbalance, normalization, and the role of Vocal Fatigue Index scores. *Applied Sciences* 2021; 11(10):4335.

[3] Nanjundeswaran C, Jacobson BH, Gartner-Schmidt J, Verdolini Abbott K. Vocal Fatigue Index (VFI): Development and validation. *J Voice* 2015; 29:433–440.

# CLINICAL VOICE BOX: DESARROLLO DE UNA APP PARA DISPOSITIVOS MÓVILES

**Rodrigo Maximiliano Jerez**[1]

[1] Facultad de Ciencias de la Salud, Universidad Católica de La Plata, La Plata, Buenos Aires, Argentina

***Palabras clave*:** clinical voice box, app, evaluación vocal, diagnóstico vocal, terapia de la voz, tecnología

## INTRODUCCIÓN

Los procesos de evaluación, diagnóstico y abordaje en el área de la voz han cambiado drásticamente en este último tiempo. La tecnología ha resultado una gran aliada permitiendo el acercamiento entre paciente y terapeuta a través de la telemedicina, sin embargo, el uso de otras herramientas como las apps ha ido cobrando relevancia y el desarrollo de recursos específicos para el área de la voz es cada vez más necesario.

El objetivo del presente trabajo es presentar el desarrollo de la App CLINICAL VOICE BOX ©, como un recurso ideado para facilitar y optimizar los procesos de evaluación, diagnóstico y abordaje en el área de la voz en entornos presenciales o de manera remota con asistencia a través de telemedicina.

## MÉTODO

A partir de una extensa revisión bibliográfica se clasificaron y agruparon los distintos instrumentos de evaluación y estrategias de abordaje. Se diseñó y programó un entorno gráfico que contenga la información, la exponga de manera amigable a los usuarios y facilite el flujo de información e interacción entre terapeuta y paciente.

## RESULTADOS

La app CLINICAL VOICE BOX permite administrar y realizar con rapidez diferentes evaluaciones y estrategias de abordaje, brinda resultados de manera automática, y almacena, recupera y compara la información en distintos momentos del tratamiento.

## DISCUSIÓN

Existen diversas aplicaciones para dispositivos móviles que pueden resultar útiles para implementar en contexto de terapia o teleterapia bien sea en sistemas operativos Android como iOS. Algunos de estos recursos permiten la obtención de ciertas medidas acústicas, otros facilitan la obtención de medidas aerodinámicas, otros más automatizados realizan evaluación perceptual según escalas específicas o permiten el entrenamiento para su aplicación en contextos reales. También se han desarrollado aplicaciones con el fin de almacenar información acerca de los tratamientos y compartirla con el paciente, aunque sin brindar mayor interacción.

La aplicación CLINICAL VOICE BOX © combina la obtención de datos en relación con diferentes pruebas e integra a la evaluación y diagnóstico, la posibilidad de prescribir una terapia, interactuar con el paciente y conocer el estado y evolución del mismo a partir de su adherencia y realización de las estrategias terapéuticas propuestas.

## CONCLUSIÓN

La app CLINICAL VOICE BOX representa un recurso tecnológico útil que facilita los procesos de evaluación, diagnóstico y abordaje en el área de la voz. Investigaciones futuras pueden centrarse en su facilidad para crear bases de datos, integrar nuevos instrumentos de examinación y comparar diferentes categorías de información a partir de programación e inteligencia artificial.

## REFERENCIAS

[1] Braun, S., Annovazzi, C., Botella, C., Bridler, R., Camussi, E., Delfino, J. P., … Stassen, H. H. (2017). Assessing Chronic Stress, Coping Skills, and Mood Disorders through Speech Analysis: A Self-Assessment “Voice App” for Laptops, Tablets, and Smartphones. *Psychopathology*, *49*(6), 406–419. https://doi.org/10.1159/000450959

[2] Bryson, P. C., Benninger, M. S., Band, J., Goetz, P., & Bowen, A. J. (2018). Telemedicine in laryngology: Remote evaluation of voice disorders-setup and initial experience. *Laryngoscope*, *128*(4), 941–943. https://doi.org/10.1002/lary.26975

[3] Kojima, T., Hasebe, K., Fujimura, S., Okanoue, Y., Kagoshima, H., Taguchi, A., … Hori, R. (2020). A New iPhone Application for Voice Quality Assessment Based on the GRBAS Scale. *Laryngoscope*, 1–3. https://doi.org/10.1002/lary.28796

[4] Kuperstock, J. E., Horný, M., & Platt, M. P. (2019). Mobile app technology is associated with improved otolaryngology resident in-service performance. *Laryngoscope*, *129*(1), E15–E20. https://doi.org/10.1002/lary.27299

[5] Munnings, A. J. (2020). The Current State and Future Possibilities of Mobile Phone “Voice Analyser” Applications, in Relation to Otorhinolaryngology. *Journal of Voice*, *34*(4), 527–532. https://doi.org/10.1016/j.jvoice.2018.12.018

[6] Oliveira, G., Fava, G., Baglione, M., & Pimpinella, M. (2017). Mobile Digital Recording: Adequacy of the iRig and iOS Device for Acoustic and Perceptual Analysis of Normal Voice. *Journal of Voice*, *31*(2), 236–242. https://doi.org/10.1016/j.jvoice.2016.05.023

[7] Petrizzo, D., & Popolo, P. S. (2020). Smartphone Use in Clinical Voice Recording and Acoustic Analysis: A Literature Review. *Journal of Voice*. https://doi.org/10.1016/j.jvoice.2019.10.006

[8] van Leer, E., Pfister, R. C., & Zhou, X. (2017). An iOS-based Cepstral Peak Prominence Application: Feasibility for Patient Practice of Resonant Voice. *Journal of Voice*, *31*(1), 131.e9-131.e16. https://doi.org/10.1016/j.jvoice.2015.11.022

# EARLY IMPACT OF COVID-19: CLINICAL VOICE PRACTICES AND LARYNGECTOMEES' EXPERIENCES

**Jeff Searl[1], Ann Kearney[2], Philip C. Doyle[2]**

[1] Department of Communicative Sciences and Disorders, Michigan State University, East Lansing, Michigan, U.S.A.
[2] Department of Otolaryngology-Head & Neck Surgery, Stanford University, Palo Alto, California, U.S.A.

***Keywords*: COVID-19, clinical practice, aerosol generating procedures, laryngectomy**

## INTRODUCTION

Because aerosolization of the SARS CoV-2 virus is a primary transmission route, clinicians performing aerosol generating procedures (AGPs) are at elevated risk of COVID-19 infection [1]. Speech-language pathologists (SLPs) providing voice, swallow, and alaryngeal care complete AGPs such as tracheoesophageal prosthesis (TEP) changes and endo-/stroboscopy. The aims of this study were to describe the early impacts of COVID-19 on 1) AGP practices by SLPs, and 2) clinical care from the total laryngectomy (TL) patient perspective, a group at elevated risk of contracting and spreading the virus [2].

## METHODS

*Participants*

Participants were 665 SLPs from Canada and the USA working in settings performing endoscopies or TEP changes and 173 TL patients >18 years old (57% TEP speech, 25% electrolarynx (EL), 9% esophageal speech (ES), and 10% writing or AAC).

*Instrumentation and Measurement*

Two questionnaires were developed. One queried SLPs about clinical impacts from COVID-19. The electronic survey (Qualtrics) was distributed via social media and other routes and was open from May 19-June5, 2020. The other asked TL patients about clinical activities, SLP advice and self-implemented changes alaryngeal speech. The survey was open from July 5-August 5, 2020.

*Analysis*

Descriptive statistics including percentages, means, standard deviations, and ranges were computed depending on the level of measurement for a survey item.

## RESULTS

Fifty percent of SLPs felt they had contracted the virus and 23% had a formal COVID-19 test with 6% positive.

**Table 6. Summary of SLP survey results.**

| Parameter | Summary |
|---|---|
| Endoscopy | Pre-COVID: 39% @ >10/wk vs 3% at survey |
| TEP changes | Pre-COVID: 24% @ >5/wk vs 6% at survey |
| PPE use | 92%-Mask with re-use by 82%; 92%-Gloves |
| Screen Patient | 81% patients pre-screened before SLP visit |

Negative financial impact occurred for 47% of the SLPs and 9% were furloughed or laid-off from their position.

**Table 2. Summary of Laryngectomee survey results.**

| Parameter | Summary | | |
|---|---|---|---|
| SLP visits and contacts | 42% 1+ in-person visit; 19% canceled visits<br>43% contacted by phone, email, text | | |
| Required Patient Precautions at SLP appointment | 77% COVID screen & 12% COVID test<br>78% mask over nose/mouth<br>73% mask over tracheostoma | | |
| Changes made to Alaryngeal Speech and Tracheostoma Coverage | | SLP Advised | Self-Implemented |
| | TEP: | 27% | 54% |
| | EL: | 12% | 45% |
| | ES: | 7% | 44% |
| | Stoma: | 21% | 47% |

## DISCUSSION AND CONCLUSION

Significant decreases in performance of endoscopy and TEP changes by SLPs occurred as anticipated. Centers for Disease Control (CDC, USA) recommendations were not always followed, including 8% of SLPs not wearing any mask and patients were not always pre-screened before an in-person visit. SLPs clinical practice, and finances were impacted with variations in severity based on work setting.

The TL patient results emphasized that in-person care is required at times even during a pandemic. CDC guidelines for COVID-19 screenings and mask wearing were not enforced for all patients for in-person visits. TL patients self-initiated changes to their alaryngeal speech more often than advised by their SLP. Changes tended to focus on stoma, neck, EL device, and hand hygiene.

Both surveys are being re-opened now one year later to gauge changes and gather data on TL patients regarding diagnoses of COVID-19 and vaccination rates.

## REFERENCES

[1] Kligerman MP, Vukkadala N, Tsang RKY, Sunwoo JB, Holsinger FC, Chan JYK, Starmer HM. Managing head and neck cancer patients with tracheostomy or laryngectomy during the COVID-19 pandemic. *Head & Neck* 2020;42(6):1209-1213.

[2] Hennessy M, Bann D, Patel VA, Saadi R, Krempl GA, Deschler DG…Choi KY. Commentary on the management of total laryngectomy patients during the COVID-10 pandemic. *Head & Neck* 2020;42(6):1137-1143.

# AUTOMATIC IDENTIFICATION OF VOICE PATHOLOGY USING DEEP NEURAL NETWORKS

**Ian S. Howard[1], Julian McGlashan[2], Evelyn Abberton[3] & Adrian J. Fourcin[4]**

[1]SECAM, University of Plymouth, Plymouth, UK.
[2]ENT Department, Nottingham University Hospitals, Nottingham, UK.
[3] University College London, UK.
[4]Laryngograph Ltd, Wallington UK, Emeritus Professor at University College London, UK.
***Keywords*: Voice pathology; Electroglottography; Deep neural networks; Telemedicine**

## INTRODUCTION

Automatic detection and preliminary classification of pathological speech signals into broad aetiological disorder categories would assist the clinical management of voice pathology. Recently much progress has been made using deep neural networks to accurately reconstruct EGG signal [1] directly from the acoustic signal [2] including from pathological speech [3], facilitating future electrode-free analytic voice-based telemedicine. Here we extend this approach to directly detect and broadly classify voice pathologies from the speech signal alone.

## METHODS

We make use of simultaneously recorded speech and EGG signals from over 100 speakers, with normal and pathological voices, reading standard passages. The pathological speakers were patients who exhibited a range of conditions; mainly adductor SD+-tremor and abductor SD, but there were also single examples of cyst, muscle tension dysphonia and cancer. We first down sampled the data signals to 16KHz and processed the EGG waveform to identify the presence of normal or potentially pathological phonation. We then train a deep neural network to directly map raw input speech and EGG waveforms to normal or clinically pre-determined pathological voice labels, obtained automatically using the EGG signal (Fig. 1 LHS). The CNN used an input window size of 1001 adjacent speech samples, had 3 convolutional layers with an input width of 20, made use of ReLu output activations and a max-pooling factor of 2. The output layer was fully connected.

## RESULTS

This new CNN provided a fine-grained indication of the presence and nature of normal and potentially pathological phonation. We present practical results of the application of network operations on normal and potentially pathological data. We show they provide a useful indication of pathology which can form the basis for classification (Fig.1 RHS). The trained CNN also responded to features in normal speech such as breathiness and irregularity, which arise in specific prosodic and discourse structures. Conversely, pathological speech generally contains regions where phonation appears normal. Taken together with phonetic context, our analysis facilitates quantitative discrimination between normal and pathological voice. This offers potential for clinical assessment without EGG.

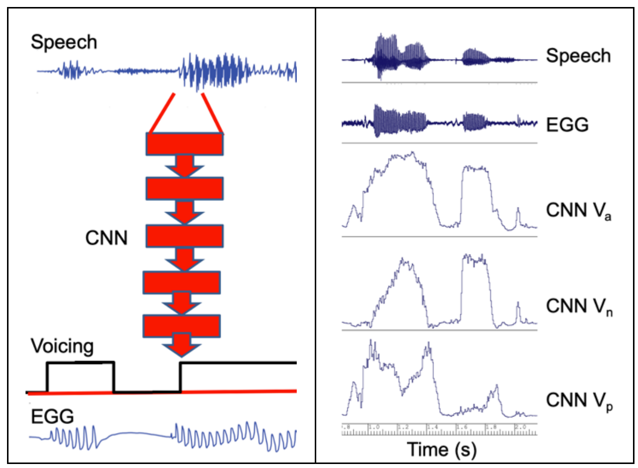



Figure 1. Left: Training CNN to estimate voice pathology by learning mapping from the speech signal alone. Right: Typical CNN output from detectors of pathological speech to all $V_a$, normal $V_n$ and pathological voicing labels $V_p$.

## ACKNOWLEDGMENTS

We thank Mark Huckvale and Reza Nouraei for useful discussion and the University of Plymouth for support.

## REFERENCES

[1] Fourcin A.J. and Abberton E, First applications of a new laryngograph, The Volta Review, 74(3), 161–176.

[2] Prathosh, A.P. Srivastava, V, & Mishra, M. (2019) Adversarial approximate inference for speech to electroglottograph conversion. IEEE/ACM Transactions on Audio, Speech, and Language.

[3] Howard I.S., McGlashan J, & Fourcin A.J. (2021) Machine learning analysis of speech and EGG for the diagnosis of voice pathology, ESSV 2021, TU Berlin, Germany.

# KINEMATIC MUCOSAL WAVE MODEL OF THE VOCAL FOLDS FOR SIMULATING KYMOGRAPHIC IMAGES

**Jan G. Švec[1], Hugo Lehoux[1], R. Sandhanakrishnan[2], S. Pravin Kumar[2]**

[1] Voice Research Laboratory, Department of Experimental Physics, Faculty of Science, Palacky University, Olomouc, Czechia
[2] Sri Sivasubramaniya Nadar College of Engineering, Department of Biomedical Engineering, Kalavakkam, Chennai, India

***Keywords*: Vocal Fold Vibration; Mucosal Waves; Modelling; Videokymography**

## INTRODUCTION

We developed a kinematic model, which mathematically simulates vocal fold (VF) oscillations with mucosal waves and produces synthetic kymograms similar to those obtained through high-speed laryngeal imaging methods in clinical practice. The purpose of this study is to briefly explain the model and demonstrate its possibilities. For a more detailed description of the model, see our previous publication [1].

## METHODS

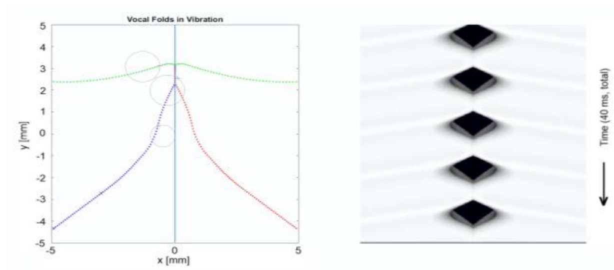
The model maps the changes of the coronal shape of the VFs through vibration cycles (Fig.1, left). The VF geometry is based on a parametrically adjustable M5 model [2].

The VF initial settings and their vibratory pattern can be adjusted by setting 17 main parameters: 1) the initial glottal halfwidth; 2,3) left and right (LR) VF thickness; 4,5) LR vertical convergence angle, 6,7) LR frequency of oscillations; 8,9) LR vertical phase differences; 10,11) LR amplitudes of the lower VF margin; 12,13) LR amplitudes of the upper VF margin; 14,15) LR mucosal wave extent on the upper VF surface; 16,17) LR initial phases of oscillations.

A kinematic rule is used for simulating the propagation of the mucosal wave from the bottom of the VF upwards and laterally over the upper VF surface [1]. The mucosal wave speed is derived from the defined vertical phase differences and the VF thickness. The vibration characteristics including the mucosal wave movements are then visualized using a synthetic kymogram graphically obtained via a virtual high-speed kymographic camera through a local illumination method [1].

## RESULTS

The model showed to be capable of simulating various types of vibratory patterns similar to those observed in clinical practice from clients with normophonic and pathologic voices. Examples of these are: soft and loud voice (controlled by the amplitude of VF vibrations), breathy and pressed voice (changing the glottal halfwidth), chest and falsetto registers (controlling frequency and vertical phase differences). Various pathologic patterns were obtained by setting different parameters for the left and right VFs.




F

Figure 1. The model: the dynamic VF shape with the trajectories of the upper and lower margins (left) and the resulting synthetic kymogram (right).

## DISCUSSION AND CONCLUSION

The model can serve as an educational and research tool for investigating the vocal fold vibratory parameters and mucosal wave features and their appearance in laryngeal kymographic images. It can also be used to estimate VF kinematic parameters from kymograms captured by high-speed or videokymographic cameras in patients with voice disorders [3].

## ACKNOWLEDGMENTS

This work has been supported by the Czech Science Foundation (GA CR) project no. 19-04477S.

## REFERENCES

[1] S. P. Kumar and J. G. Svec, "Kinematic model for simulating mucosal wave phenomena on vocal folds," *Biomed. Signal Process. Control,* vol. 49, pp. 328-337, 2019.

[2] 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,* vol. 109, no. 4, pp. 1616-1630, 2001.

[3] S. Bulusu, S. P. Kumar, J. G. Svec, and P. Aichinger, "Fitting synthetic to clinical kymographic images for deriving kinematic vocal fold parameters: Application to left-right vibratory phase differences," *Biomed. Signal Process. Control,* vol. 63, 2021, Art no. 102253, doi: 10.1016/j.bspc.2020.102253.

*The 14th International Conference on Advances in Quantitative Laryngology, Voice and Speech Research (AQL)*

# NEURAL-NETWORK-BASED ESTIMATION OF VOCAL FOLD KINEMATIC PARAMETERS FROM DIGITAL VIDEOKYMOGRAMS

**S. Bulusu[1], S. P. Kumar[2], J. G. Švec[3], P. Aichinger[1]**

[1] Medical University of Vienna, Department of Otorhinolaryngology, Division of Phoniatrics-Logopedics, Vienna, Austria

[2] Sri Sivasubramaniya Nadar College of Engineering, Department of Biomedical Engineering, Kalavakkam, Chennai, India

[3] Palacky University, Faculty of Science, Department of Experimental Physics, Voice Research Laboratory, Olomouc, Czech Republic

***Keywords*: High-Speed Videoendoscopy, Machine Learning, Convolutional Neural Network, Dysphonia**

## INTRODUCTION

Kinematic parameters of vocal fold (VF) vibration may help revealing the vocal health status. It was shown previously that kinematic parameters can be estimated using a digital kymogram (DKG) simulator combined with golden section search [1,2]. However, the search is computationally expensive. This work concerns the use of Convolutional Neural Networks (CNNs) for estimating kinematic VF parameters from input DKGs.

## METHODS

172 laryngeal high-speed videos of 43 healthy subjects and 49 patients with a wide variety of different voice disorders were used [3]. First, a total of 163.656 DKGs were extracted in a semi-automatic way. Second, the ground truth kinematic parameters were extracted using the previously published search method [2]. Third, one separate CNN similar to the classic LeNet was trained for each of the kinematic parameters. The training set consisted of 100.000 images, the validation set of 50.000, and the test set of 13.656.

## RESULTS

Table 1 shows the RMSE of the kinematic parameters output by the CNNs, the parameter ranges, and the relative RMSE. The ten parameters are the frequency $f_o$, the glottal half width $w_h$, the upper and lower VF margins' amplitudes $A_u$ and $A_l$, the vertical and left-to-right phase differences $\Phi_{vertical}$ and $\Phi_{LR}$, the divergence angle α, common phase $\Phi_{comm}$, the VF thickness T, and the mucosal wave extent E.

## DISCUSSION AND CONCLUSION

The presented results suggest that the proposed CNN-based regression is competitive to the search method proposed previously. The CNNs are faster (fractions of seconds versus hours), while their RMSE accuracy is comparable. A residual mismatch between the used VF vibration model and the natural data may limit the accuracy of kinematic parameter regression.

**Table 7. Performance of the CNNs for each parameter by means of absolute and relative root-mean-square-errors (RMSE).**

| **Parameter** | **$f_o$ (Hz)** | **$\Phi_{LR}$ (rad)** | **$\Phi_{comm}$ (rad)** | **$A_l$ (cm)** | **$A_u$ (cm)** |
|---|---|---|---|---|---|
| **RMSE** | 7.97 | 0.8 | 0.83 | 0.016 | 0.0032 |
| **Range** | 70-300 | -3.14-3.14 | -3.14-3.14 | 0.02-0.1 | 0.05-0.1 |
| **RMSE %** | 3.47% | 12.7% | 13.2% | 20% | 6.4% |
| **Parameter** | **$\Phi_{vertical}$ (°)** | **$w_h$ (cm)** | **α (°)** | **T (cm)** | **E** |
| **RMSE** | 8.87 | 0.0053 | 5.66 | 0.048 | 26.37 |
| **Range** | 10-120 | 0-0.15 | -10-40 | 0.2-0.5 | 0.25-100 |
| **RMSE %** | 8.06% | 3.53% | 11.32% | 16% | 26.44% |

## ACKNOWLEDGMENTS

This work was supported by the Austrian Science Fund (FWF): KLI 722-B30, and the Czech Science Foundation (GA CR) project no. 19-04477S.

## REFERENCES

[1] S.P. Kumar, J.G. Švec, Kinematic model for simulating mucosal wave phenomena on vocal folds, Biomed. Signal Process. Control. 49 (2019) 328–337. https://doi.org/10.1016/j.bspc.2018.12.002.

[2] S. Bulusu, P.K. Subbaraj, J.G. Svec, P. Aichinger, Fitting synthetic to clinical kymographic images for deriving kinematic vocal fold parameters: Application to left-right vibratory phase differences, Biomed. Signal Process. Control. 63 (2021) 102253. https://doi.org/10.1016/j.bspc.2020.102253.

[3] P. Aichinger, I. Roesner, M. Leonhard, D.-M. Denk-Linnert, W. Bigenzahn, B. Schneider-Stickler, A Database of Laryngeal High-Speed Videos with Simultaneous High-Quality Audio Recordings of Pathological and Non-Pathological Voices, in: Proc. Tenth Int. Conf. Lang. Resour. Eval., European Language Resources Association (ELRA), Portorož, Slovenia, 2016: pp. 767–770. https://www.aclweb.org/anthology/L16-1122.

# PARAMETER ANALYSIS AND UNCERTAINTIES OF IMPEDANCE-BASED INVERSE FILTERING FROM NECK SURFACE ACCELERATION

**R. Manríquez[1], V. M. Espinoza[2], C. Castro[2,3], J. P. Cortés[4], M. Zañartu[1]**

[1] Department of Electronic Engineering, Universidad Técnica Federico Santa María, Valparaíso, Chile
[2] Department of Sound, Universidad de Chile, Santiago, Chile
[3] Department of Communication Disorders, University de Valparaíso, Valparaíso, Chile
[4] Center for Laryngeal Surgery and Voice Rehabilitation Laboratory, Massachusetts General Hospital, Boston, Massachusetts

***Keywords*: Accelerometer, glottal airflow, neck vibration, voice production**

## INTRODUCTION

The purpose of this study is to determine the uncertainty in the use of an Impedance-Based Inverse Filtering (IBIF) method that estimates glottal airflow from neck-surface acceleration. The level of uncertainty is evaluated in terms of how the filter parameters behave and how these variations propagate into the estimated measures obtained from the signal.

## METHODS

### *Participants*

Two groups of adult female individuals were considered: 19 individuals (aged 25.2 ± 2.7 years) with phonotraumatic vocal hyperfunction (PVH) and 19 healthy control participants (aged 26.3 ± 2.4 years).

### *Instrumentation and Measurement*

The data acquisition protocol was based on methods from previous studies [1]. The oral airflow volume velocity (OVV) and the neck-surface accelerometer signal (ACC) were simultaneously acquired during five repetitions of two phonemes /a/ and /i/.

### *Analysis*

Inverse filtering of the OVV signal was accomplished by removing the effect of the first formant (F1) by means of a single notch filter (SNF), following methods presented in previous related studies [1].

The IBIF model requires a calibration step to obtain approximations of glottal airflow from the neck-skin ACC signal. The procedure is reported by Zañartu et al [2]. An additional gain parameter G is included in this work to compensate for acceleration units. Subject-specific Q parameters for the IBIF model were determined to minimize the waveform error between the OVV-based glottal airflow (reference signal) and the inverse-filtered neck-skin ACC signal (signal to be matched to the reference signal).

## RESULTS

Significant differences between Q values for /a/ and /i/ are observed ($p<0.01$). This is observed in both groups. Regarding aerodynamic measures, for AC Flow relative error for vowel /a/ was of 12.9% and 18.1% for healthy and PVH groups, respectively. For /i/, error was below 7.8% for both groups. For MFDR, error was below 10% for /a/ and below 14.1% for /i/ in both groups.

A cross-estimation method using the opposite filter values instead of the corresponding ones is tested. It is observed that here, estimation errors drastically increase. For AC Flow, a 20%> error is obtained for the /a/ vowel and a 40%-50%> error for the /i/. For MFDR, error ranges from 35-60\% for both vowels.

To compensate the increasing error, an alternative method to choose Q values is proposed: using weighted Q values chosen from both vowels. With this method, it is found that an error below 20% for both measures in all groups and vowels can be obtained, except for MFDR for the /a/ in which an error below 26% is guaranteed.

## DISCUSSION AND CONCLUSIONS

From these results, it can be concluded that estimation of aerodynamic measures is greatly affected by the Q values chosen. Error increases when performing a cross-estimation using the opposite filter values, but a compromise can be found between both vowels using a weighted method.

## 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] Espinoza VM, Zañartu M, Van Stan JH, Mehta DD, and Hillman RE, "Glottal aerodynamic measures in women with phonotraumatic and nonphonotraumatic vocal hyperfunction", *J Speech Lang Hear Res*, vol. 60(8), pp. 2159–2169, 2017.

[2] Zañartu M, Ho JC, Mehta DD, Hillman RE, and Wodicka GR, "Subglottal Impedance-Based Inverse Filtering of Voiced Sounds Using Neck Surface Acceleration's Audio, Speech, and Language Processing, IEEE Transactions on, vol. 21, no. 9, pp. 1929–1939, Sep. 2013.

# PHONOSIM: A PATIENT-SPECIFIC COMPUTATIONAL MODELING SUITE FOR PHONOSURGERY

**Zheng Li[1], Azure Wilson[2], Lea Sayce[2], Amit Avhad[1], Bernard Rousseau[2], Haoxiang Luo[1]**

[1] Department of Mechanical Engineering, Vanderbilt University, Nashville, TN, USA

[2] Department of Communication Science and Disorders, University of Pittsburgh, Pittsburgh, PA, USA

***Keywords*: Fluid-structure interaction; Unilateral vocal fold paralysis; Patient specific modelling; Vocal fold vibration**

## INTRODUCTION

Unilateral vocal fold paralysis (UVFP) is a deliberating voice disorder, resulting in impaired adduction, one-sided vocal fold immobility, and loss of voice. We are developing a novel computational modelling suite for the investigation of UVFP, which will be used to inform future in silico approaches to improve surgical outcomes in type I thyroplasty.

## METHODS

Our cross-disciplinary team has developed a suite of computational tools to simulate the fluid-structure interaction (FSI) of vocal fold vibration, which is coined "PhonoSim" for phonosurgery simulation. The suite includes finite-element method (FEM) models of the larynx, simplified FSI models of the vocal fold vibration, and high-fidelity FSI models that resolve both 3D glottal airflow and the tissue deformation. Furthermore, all of these models are based on the patient-specific anatomy as shown in Figure 1, where 3D image data from MRI scan are used to construct the FEM tissue model.

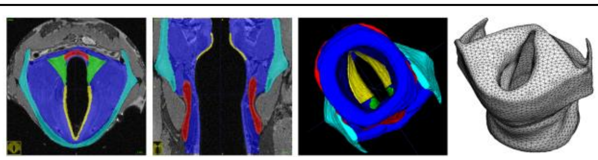


Figure 1. FEM model development from MRI scan

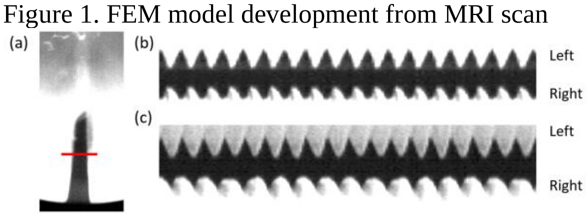



Figure 2. HSV tests for healthy phonation and UVFP

In the experiments, a surgical approach is developed using a rabbit model to simulate both the healthy phonation and the UVFP vibration *in vivo* [1, 2]. High-speed videos (HSV) of vocal fold vibrations are recorded for different conditions in phonation tests as shown in Figure 2, which are essential to study UVFP and critical for numerical model validation.

## RESULTS AND DISCUSSION

PhonoSim can perform different analyses of vocal fold vibration as shown in Figure 3, collectively, these models of different complexities enhance the overall modeling accuracy and efficiency for individual patients. With FEM model, eigenfrequencies of two sides of vocal fold agree well with each other in health phonation and show valid differences in the UVFP configuration [2]. With this FEM model, the initial implant shape and position can be determined. A 1D unsteady flow model enhanced by machine learning has been developed to couple with the 3D FEM model for FSI simulation to further estimate the tissue properties [3] and to optimize the implant. The high-fidelity, full-3D FSI model is used to verify and finalize the implant design.

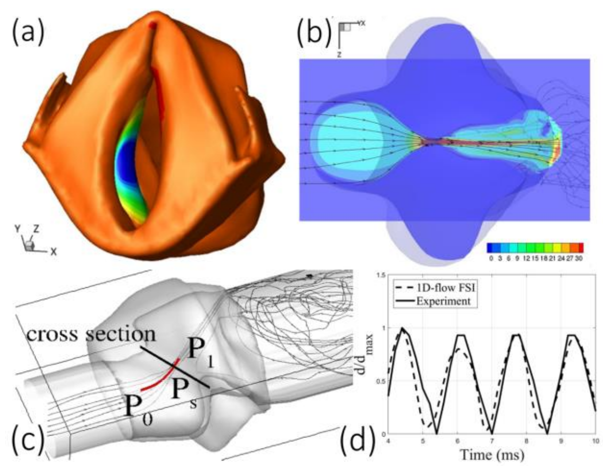



Figure 3. Complementary modeling analyses from PhonoSim: (a) Eigenmode from FEM model; (b) 3D FSI; (c) 1D flow model for glottal airflow; (d) vibration from 1D flow FSI.

## ACKNOWLEDGMENTS

This research was supported by NIH grant funding NIDCD under Award Number 5 R01 DC016236 03 and the NCATS under Award Number TL1TR001858.

## REFERENCES

[1] Chang et al. "Subject-specific computational modeling of evoked rabbit phonation." Journal of biomechanical engineering 138.1 (2016).

[2] Li et al. "Numerical and Experimental Investigations on Vocal Fold Approximation in Healthy and Simulated Unilateral Vocal Fold Paralysis." Applied Sciences 11.4 (2021): 1817.

[3] Li et al. "A one-dimensional flow model enhanced by machine learning for simulation of vocal fold vibration." The Journal of the Acoustical Society of America 149.3 (2021): 1712-1723.

*The 14th International Conference on Advances in Quantitative Laryngology, Voice and Speech Research (AQL)*

# *SIMVOICE* – PARAMETER STUDY ON GLOTTAL INSUFFICIENCY AND APERIODIC VOCAL FOLD OSCILLATIONS

**Sebastian Falk[1], Stefan Kniesburges[1], Bernhard Jakubaß[1], Matthias Echternach[3], Paul Maurerlehner[2], Stefan Schoder[2], Manfred Kaltenbacher[2], Anne Schützenberger[1], Michael Döllinger[1]**

[1] Div. of Phoniatrics and Pediatric Audiology, Dept. of Otorhinolaryngology, University Hospital Erlangen, Germany
[2] Inst. of Fundamentals and Theory in Electrical Eng., Div. Vibro- and Aeroacoustics, Graz University of Technology, Austria
[3] Div. of Phoniatrics and Pediatric Audiology, Dept. of Otorhinolaryngology, Munich University Hospital (LMU), Germany

***Keywords*: comp. fluid dynamics (CFD), comp. aeroacoustics (CAA), glottal insufficiency, aperiodicity**

## INTRODUCTION

The central aim of the *simVoice* project is to develop a three-dimensional aero-acoustic numerical larynx model for future application in clinical environment. After validating *simVoice* toward an experimental setup [1, 2], we now investigate the ability of *simVoice* to mimic symptoms of functional voice disorders as glottal insufficiency and aperiodic vocal fold oscillations.

The phonation process is assumed to be most efficient when (1) the vocal folds close the glottal gap completely in each oscillation and (2) oscillate symmetrically and periodically. A glottal insufficiency and an aperiodic oscillation of the vocal folds cause a reduced voice quality. Numerical models based on Finite-Elements (FE) and/or Finite-Volumes (FV) have great potential to investigate the cause-and-effect-chain of a disturbed phonation process.

## METHODS

*simVoice* is a hybrid model. It consists of a fluid dynamic simulation model with external driven vocal fold motions, based on the 3D FV method [1], and an aero-acoustic model, based on the 3D FE method [2]. The numerical larynx model simVoice considers the fluid flow through the glottis, the vocal fold (VF) motions, and the resulting acoustic signal. In this study, four types of clinically seen glottis closure types (GC1-GC4) were modeled [3] that are based on high-speed recordings obtained from experiments with ex vivo porcine larynges [4]. GC1-GC4 represent posterior gaps with an increasing glottal insufficiency. In addition to the previously developed periodic (symmetric and asymmetric) vocal fold motion [3], aperiodic oscillations (symmetric and asymmetric) based on in vivo recordings were modeled. Furthermore, to mimic soft/quiet, normal, and loud voice, we chose 385 Pa, 775 Pa, and 1500 Pa as subglottal pressure. Overall, combining all motion patterns, GC types, and subglottal pressures, we conducted 48 simulations.

## RESULTS

Our results show that the mean glottis' volume flow increases with increasing gap and increasing subglottal pressure but decreases with asymmetric or aperiodic vocal fold oscillations. For phonation, the goal is to increase the energy transfer between the airflow and the VFs. The total transferred work ($W_{net}$) during one oscillation cycle for all GC types and motion cases is positive but decreases with an increasing glottal insufficiency. A periodic and symmetric motion of the vocal folds results in a higher $W_{net}$ than that of periodic asymmetric motion, whereas the aperiodic oscillation decreases $W_{net}$. Furthermore, $W_{net}$ is significantly increased by increasing the subglottal pressure. The spectra-based and well-established Cepstral Peak Prominence (CPP) parameter decreases for an increasing glottal insufficiency. The CPP for a symmetric and periodic oscillation combined with a high subglottal pressure reaches the highest values. Asymmetry and aperiodicity decrease CPP.

## CONCLUSION

Our study shows that a high subglottal pressure combined with a fully or partially closed glottis (GC1-GC3) and a symmetric and periodic vocal fold oscillation achieves the highest quality of the acoustic signal. *simVoice* confirms previous clinical and experimental observations that a high level of glottal insufficiency worsens the acoustic signal quality more than an asymmetric or aperiodic oscillation. All symptoms combined further reduce the quality of the sound signal.

## ACKNOWLEDGMENTS

The authors acknowledge support from the German Research Foundation (DFG) under DO 1247/10-1 (no. 391215328) and the Austrian Research Council (FWF) under no. I 3702.

## REFERENCES

[1] H. Sadeghi, S. Kniesburges, M. Kaltenbacher, A. Schützenberger, and M. Döllinger, *Computational models of laryngeal aerodynamics: Potentials and numerical costs*. Journal of Voice, 2018.

[2] 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., 2020.

[3] 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.* Frontiers in Physiology, 2021.

[4] 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., 2017.

# AUTOMATED DETECTION AND SEGMENTATION OF GLOTTAL AREA USING DEEP-LEARNING NEURAL NETWORKS IN HIGH-SPEED VIDEOENDOSCOPY DURING CONNECTED SPEECH

**Ahmed M. Yousef[1], Dimitar D. Deliyski[1], Stephanie R.C. Zacharias[2,3], Alessandro de Alarcon[4,5], Robert F. Orlikoff[6], Maryam Naghibolhosseini[1]**

[1] Department of Communicative Sciences and Disorders, Michigan State University, East Lansing, Michigan, USA
[2] Mayo Clinic-Arizona, Scottsdale, Arizona, USA
[3] Department of Otolaryngology-Head and Neck Surgery, Mayo Clinic, Phoenix, Arizona, USA
[4] Division of Pediatric Otolaryngology, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, USA
[5] Department of Otolaryngology Head and Neck Surgery, University of Cincinnati, Ohio, USA
[6] College of Allied Health Sciences, East Carolina University, Greenville, North Carolina, USA

***Keywords*: High-Speed Videoendoscopy; Connected Speech; Spatial Segmentation; Deep Learning**

## INTRODUCTION

Voice disorders typically reveal themselves not in sustained phonation, but in connected speech. Although videostroboscopy is used for clinical voice assessment in connected speech, without a sufficient temporal resolution, it fails to provide intra-cycle details of vocal fold (VF) vibration [1]. Laryngeal high-speed videoendoscopy (HSV) overcomes this limitation, allowing the detailed imaging of the true VF vibrations in running speech [2]. Despite this, the use of HSV remains a daunting task for clinicians, as they must visually navigate through thousands of HSV frames. Therefore, it is crucial to develop automated tools to analyze VF vibration, such as segmenting VF edges.

Several image processing and machine learning methods were used to segment VF edges in HSV in isolated vowels, but not in running speech [3]. We have developed an automated approach to segment VF edges in HSV in running speech based on an active contour modeling approach [4]. We then enhanced this method by coupling it with a machine-learning technique and designed a hybrid approach [5]. Although this method was accurate, it required a high computational cost and was designed to extract the glottal edges during only VF vibrations. The current work builds upon the hybrid method and presents a more robust and less time-consuming scheme using deep learning. This method can detect and extract the glottal edges during all HSV frames in connected speech – including nonstationary events such as onsets/offsets of phonation and when the vocal folds are not vibrating.

## METHODS

### *Participants*

A vocally normal 38-year-old female was examined at the Center for Pediatric Voice Disorders, Cincinnati Children's Hospital Medical Center; the examination was approved by the Institutional Review Board.

### *Instrumentation and Measurement*

A custom HSV system recorded the subject when reading the "Rainbow Passage." A FASTCAM SA-Z color high-speed camera (Photron Inc., San Die-go, CA) with a 12-bit color image sensor and 64 GB of cache memory was coupled with a flexible fiberoptic nasolaryngoscope (3.6-mm Olympus ENF-GP Fiber Rhinolaryngoscope; Olympus Corporation, Tokyo, Japan) and a 300-W xenon light source. The recording took 29.14 s (116,543 video frames) and was obtained with 4,000 frames/second, a spatial resolution of 256x256 pixels, and 249 µs integration time. The recorded HSV data was saved as an uncompressed 24-bit RGB AVI file.

### *Analysis*

The hybrid approach we have recently developed [5] served as an automated labelling tool to build a training dataset rather than using manual labeling. The labelling tool comprised several image processing steps (see [5]): i) temporal segmentation was first performed to automatically determine the vocalized segments in the video, ii) a motion compensation was used to capture VF locations, iii) HSV kymograms were extracted at various VF cross sections, iv) a k-means clustering algorithm was utilized to segment the glottal edges in the kymograms, (v) the segmented edges were modeled as active contours that were deformed until precise capturing of the glottal edges. These segmented frames were used as fully automatically labelled data to train a deep neural network.

The training dataset contained 2050 frames and 20% of the dataset was used for validation. A fully convolutional neural network with a u-shaped architecture (U-net architecture) was employed, comprising encoder and decoder [6]. The network was trained with a batch size of 10 for maximal 20 epochs using Adam optimizer. Due to laryngeal movements in running speech, data augmentation was applied to the training data to improve the generalizability of the trained network. After training, the network was tested on manually segmented frames (471 frames), selected randomly and not included in the training dataset. Intersection-over-Union (*IoU*) metric was used as an evaluation metric:

$$IoU = TP / (TP + FP + FN), \quad (1)$$

where TP was correctly classified pixels, FN and FP were incorrectly classified pixels as nonglottal and glottal pixels, respectively.

### RESULTS

The automated labelling tool was able to segment the glottal area in 2050 HSV frames forming a training dataset. The deep neural network was successfully trained on the automatically segmented frames in the training dataset. The trained network was then tested on manually labeled HSV frames yielding a mean IoU of 0.8 in segmenting the glottal region. Fig. 1 (a) shows the performance of applying the developed network to 6 testing frames. Each frame is shown before and after segmenting the glottal area along with the associated binary segmentation mask.

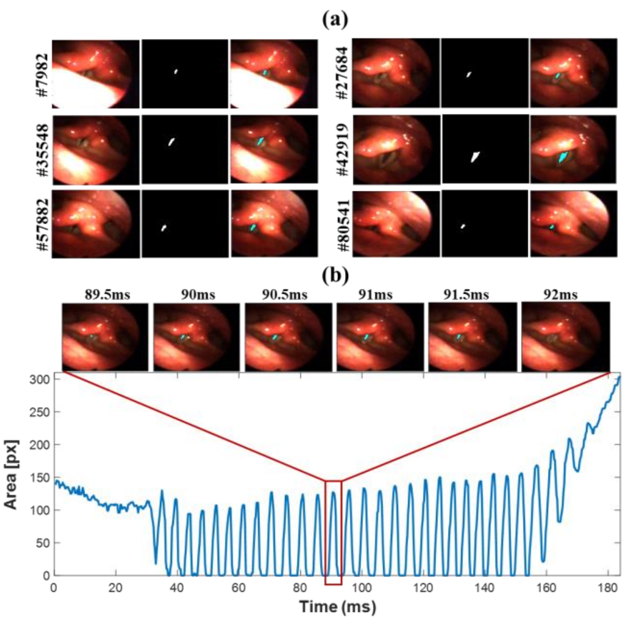



Figure 1. (a) Segmented frames using the trained network; (b) The glottal area waveform of a vocalization using the trained network (frame #109,648-110,383).

Results of extracting the glottal area waveform using the network for a vocalized segment (frame #109,648-110,383) is shown in Fig. 1 (b). Six segmented frames are also displayed with the corresponding time instances. It should be noted that this algorithm provides the glottal edge representations for the right and left vocal folds.

### DISCUSSION AND CONCLUSION

This study demonstrated the feasibility of using our previously developed image segmentation approach as an automated labelling tool. This tool provided a great advantage over manual labeling in terms of creating a large-size training dataset in a cost-effective manner, which is favorable for training machine-learning approaches. The automated labelling tool was utilized to train a deep-learning network on segmenting the glottal area in connected speech. The trained network showed a promising performance against a manually labeled dataset through accurate detection of the glottal edges and computation of the glottal area waveform. The developed approach outperformed our prior method by capturing the glottal edges during VF vibration, voicing onsets/offsets and breaks. Developing automated methods for spatial segmentation of HSV data is crucial for automated analysis of vocal function. Our developed deep-learning scheme serves as a powerful tool in analyzing VF vibration in connected speech. This tool will be used for glottal edge representation in patients with spasmodic dysphonia in future. This is toward development of HSV-based measures for voice evaluation in connected speech.

### ACKNOWLEDGMENTS

This research was funded by the National Institutes of Health, NIDCD, K01DC017751 “Studying the Laryngeal Mechanisms Underlying Dysphonia in Connected Speech” and NIDCD R01-DC007640 “Efficacy of Laryngeal High-Speed Videoendoscopy.”

### REFERENCES

[1] Zacharias, S. R., Myer IV, C. M., Meinzen-Derr, J., Kelchner, L., Deliyski, D. D., & de Alarcón, A. (2016). Comparison of videostroboscopy and high-speed videoendoscopy in evaluation of supraglottic phonation. Annals of Otology, Rhinology & Laryngology, 125(10), 829-837.

[2] Naghibolhosseini, M., Deliyski, D. D., Zacharias, S. R., de Alarcon, A., & Orlikoff, R. F. (2018). Temporal segmentation for laryngeal high-speed videoendoscopy in connected speech. Journal of Voice, 32(2), 256-e1.

[3] Kist, A. M., & Döllinger, M. (2020). Efficient biomedical image segmentation on edgetpus at point of care. IEEE Access, 8, 139356-139366.

[4] Yousef, A. M., Deliyski, D. D., Zacharias, S. R., de Alarcon, A., Orlikoff, R. F., & Naghibolhosseini, M. (2020). Spatial Segmentation for Laryngeal High-Speed Videoendoscopy in Connected Speech. Journal of Voice.

[5] Yousef, A. M., Deliyski, D. D., Zacharias, S. R., de Alarcon, A., Orlikoff, R. F., & Naghibolhosseini, M. (2021). A Hybrid Machine-Learning-Based Method for Analytic Representation of the Vocal Fold Edges during Connected Speech. Applied Sciences, 11(3), 1179.

[6] Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.

# VISCOELASTICITY OF HUMAN LARYNGEAL MUCUS FROM THE VOCAL FOLDS

**Gregor Peters[1], Olaf Wendler[1], David Böhringer[2], Antoniu-Oreste Gostian[3], Sarina K. Müller[3], Herbert Canziani[4], Nicolas Hesse[4], Marion Semmler[1], David A. Berry[5], Stefan Kniesburges[1], Wolfgang Peukert[4], Michael Döllinger[1]**

[1] Div. of Phoniatrics and Pediatric Audiology, Dept. of Otorhinolaryngology, University Hospital Erlangen, Germany
[2] Dept. of Physics, Biophysics Group, FAU Erlangen-Nürnberg, Erlangen, Germany
[3] Dept. of Otorhinolaryngology, University Hospital Erlangen, FAU Erlangen-Nürnberg, Germany
[4] Dept. of Chemical and Biological Engineering, Chair of Particle Technology, FAU Erlangen-Nürnberg, Erlangen, Germany
[5] Dept. of Head and Neck Surgery, David Geffen School of Medicine at UCLA, Los Angeles, California, USA

***Keywords*: Human Laryngeal Mucus, Rheology, Viscoelasticity, Vocal Folds**

## INTRODUCTION

Several clinical studies reported mucus of varying thickness for persons with and without voice disorders. Specific diseases like cystic fibrosis affect mucus consistency and promote voice disorders [1]. The effects of artificial mucus on vibrational characteristics of vocal folds were already investigated [2], but without consideration of realistic viscoelastic conditions. The central objective of this study is the rheological characterization of human laryngeal mucus. The intended use of the findings is the creation of a synthetic mucus for analyzing its characteristics in dynamic ex-vivo larynx experiments. These experiments offer conditions for investigations on the effect of mucus on the oscillatory parameters of the vocal folds and will contribute to a better understanding of the importance of mucus for the phonatory process.

## METHODS

Human laryngeal mucus was gained directly from the vocal folds of patients during surgeries under general anesthesia by two surgeons. To deal with small sample amounts, particle tracking microrheology (PTM) was applied to determine mucus viscoelasticity. 19 mucus samples were evaluated, and the results were compared to five additional samples with a sufficient volume for measurement by oscillatory shear rheology (OSR). Moreover, the results were related to the demographic data.

## RESULTS

Human laryngeal mucus revealed viscoelastic diversity. Differences were found according to the rigidity of the mucus samples with both methods. The viscoelasticity of mucus obtained by OSR matched the results of PTM. Samples that were applied to PTM revealed either throughout solid-like character or a transition from solid-like to liquid-like over frequency. This led to a subdivision of the mucus samples into three groups. No relation of the groups was found for gender, age, and larynx healthiness. Nevertheless, smokers showed predominantly mucus with lower rigidity and a change from solid-like to liquid-like at lower frequencies.

## DISCUSSION AND CONCLUSION

The viscoelasticity of human laryngeal mucus varied from patient to patient. However, the measurement results obtained by PTM could be grouped due to similar rheological characteristics. The differences may be caused by variations of the hydration of the mucus, affecting the concentration of mucins, the main gel-building component, and corresponding network-building factors. It can be expected that this is governed by scaling laws as found for living cells [3], which would facilitate the creation of highly adaptable and controllable mucus substitutes.

## ACKNOWLEDGMENTS

This work was supported by the Deutsche Forschungsgemeinschaft (DFG) under grant DO1247/6-2 (no. 281313362). David Böhringer's effort on this project was supported by the Emerging Fields Initiative of the University of Erlangen-Nürnberg. Dr. Berry's effort on this project was supported by grant number R01 DC013323 from the National Institute of Deafness and Other Communication Disorders (NIDCD/NIH).

## REFERENCES

[1] Lourenço BM, Costa KM, Silva Filho M. *Voice Disorder in Cystic Fibrosis Patients.* PLOS ONE, vol. 9, pp. 1-8, 5 2014.

[2] Döllinger M, Gröhn F, Berry DA, Eysholdt U, Luegmair G. *Preliminary results on the influence of engineered artificial mucus layer on phonation.* Journal of Speech, Language, and Hearing Research, vol. 57, pp. S637-S647, 2014.

[3] Fabry B, Maksym GN, Butler JP, Glogauer M, Navajas D, Fredberg JJ. *Scaling the microrheology of living cells*. Physical Review Letters, vol. 87, pp. 148102/1-148102/4, 2001.

# RELATIONSHIPS BETWEEN THE NECK-SURFACE ACCELERATION PARAMETERS OF THE DAILY PHONOTRAUMA INDEX AND GLOTTAL AERODYNAMIC MEASURES IN VOCALLY HEALTHY FEMALES

**Juan P. Cortés[1], Daryush D. Mehta[1,2,3], Andrew J. Ortiz[1], Laura E. Toles[1,2], Katherine L. Marks[1,2], Jarrad H. Van Stan[1,2,3], Robert E Hillman[1,2,3]**

[1] Center for Laryngeal Surgery and Voice Rehabilitation, Massachusetts General Hospital, Boston, MA, U.S.A.
[2] MGH Institute of Health Professions, Massachusetts General Hospital, Boston, MA, U.S.A.
[3] Harvard Medical School, Harvard University, Cambridge, MA, U.S.A.

***Keywords*: Daily Phonotraumatic Index; Glottal Aerodynamics; Neck-Skin Acceleration**

## INTRODUCTION

The Daily Phonotraumatic Index (DPI) [1] is a recently proposed logistic regression-based index that incorporates two long-term distributional measures extracted from ambulatory recordings of neck-skin acceleration (ACC) signals (neck-surface acceleration magnitude (NSAM) skewness and standard deviation of H1-H2) to differentiate normal and phonotraumatic (e.g., vocal nodules) vocal function. The goal of this work is to help determine the validity of interpreting the DPI as reflecting the increases in laryngeal forces and abruptness of vocal fold closure that are associated with vocal fold tissue trauma. This was done by correlating the NSAM and H1-H2 parameters with glottal aerodynamic measures of vocal fold vibratory function.

## METHODS

Synchronous recordings of ACC and oral airflow were obtained as 18 vocally healthy individuals produced multiple strings of /pae/ syllables while decreasing from loud-to-soft voice. ACC features that make up the DPI include the neck-surface acceleration magnitude (NSAM; associated with overall laryngeal forces) and the difference in magnitude of the first and second harmonics (H1-H2; an indicator of the degree of glottal closure). Measures of peak-to-peak flow (AC flow) and maximum flow declination rate (MFDR) were extracted from the inverse-filtered oral airflow. Pairwise and multiple linear regressions were applied on a group and per-individual basis to assess the strength of the relationships between the DPI features (NSAM and H-H2) and glottal aerodynamic measures (AC flow and MFDR).

## RESULTS

Results for pairwise correlations showed that glottal aerodynamic measures were more strongly correlated with NSAM for the group ($r$ = 0.83 for MFDR; $r$ = 0.85 for AC flow) and individual subjects (18/18 subjects had r > 0.71 for both measures) than was the case for H1-H2 for the group ($r$ = 0.51 for MFDR; $r$ = 0.53 for AC flow) and individual (15/18 subjects had r > 0.71 for both measures). Group-based multiple linear regression (NASM + H1-H2) produced slightly higher correlations. (r = 0.84 for MFDR, 0.86 for AC flow).

## DISCUSSION AND CONCLUSION

As expected, intensity-related measures of glottal flow (AC flow and MFDR) have strong correlations with NSAM for individual participants. The differences in regression lines between subjects are mostly due to calibration differences between the ACC and oral airflow signals. The linear relationship between H1-H2 and AC flow/MFDR varies substantially across subjects and has minimum effect in a multiple regression model. This emphasizes the importance of long-term statistics (i.e., the standard deviation of H1-H2), instead of point-estimates, for the analysis of vocal fold closure related to phonotrauma.

The strong correlations between DPI parameters and glottal aerodynamic measures associated with vocal fold vibratory function support further development of DPI as an indicator of the potential risk for phonotrauma.

## ACKNOWLEDGMENTS

This work was supported by the National Institute on Deafness and Other Communication Disorders Grants R21 DC015877 (awarded to Daryush D. Mehta) and P50 DC015446 (awarded to Robert E. Hillman). The paper's contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH. Disclosure: Drs. Robert Hillman and Daryush Mehta have a financial interest in InnoVoyce LLC, a company focused on developing and commercializing technologies for the prevention, diagnosis, and treatment of voice-related disorders.

## REFERENCES

[1] Van Stan et al. Changes in a Daily Phonotrauma Index After Laryngeal Surgery and Voice Therapy: Implications for the Role of Daily Voice Use in the Etiology and Pathophysiology of Phonotraumatic Vocal Hyperfunction. *J Speech, Language, and Hearing Research* 2020; 63(12):3934:3944.

# A MACHINE LEARNING FRAMEWORK FOR ESTIMATING SUBGLOTTAL PRESSURE DURING RUNNING SPEECH FROM GLOTTAL AIRFLOW MEASURES

**Emiro Ibarra[1,2], Jesús Parra[1], Gabriel Alzamendi[3], Juan P. Cortés[4], Víctor Espinoza[5], Daryush D. Mehta[4], Matías Zañartu[1]**

[1] Department of Electronic Engineering, Universidad Técnica Federico Santa María, Valparaíso, Chile
[2] School of Electrical Engineering, Universidad de Los Andes, Mérida, Venezuela
[3] Institute for Research and Development on Bioengineering and Bioinformatics, CONICET-UNER, Entre Rios, Argentina
[4] Center for Lanryngeal Surgery and Voice Rehabilitation Laboratory, Massachusetts General Hospital, MA, USA
[5] Department of Sound, Universidad de Chile, Santiago, Chile

***Keywords*: Subglottal pressure estimation, voice production model, neural networks, clinical voice assessment**

## INTRODUCTION

The assessment of vocal function is relevant for understanding the underlying glottal mechanisms involved in healthy and disordered phonation. Recent efforts assess vocal function through an inverse problem formulation built around biomechanical models of the vocal folds in a Bayesian framework [1]. However, this method requires multimodal voice-related signals and a computational load that can hinder its applicability in an ambulatory scenario. Extending previous efforts [2-3] for in vivo scenarios, a supervised machine learning framework is proposed to construct a non-linear mapping to estimate subglottal pressure from selected aerodynamic measures during running speech. Only synthetic signals obtained from lumped-element voice production model are used in the training stage.

## METHODS

### *Instrumentation and Measurement*

The data consisted of simultaneous recordings of oral volume velocity (OVV) and intraoral pressure (IOP) corresponding to a repetitive /pa/ gesture from a male participant having no medical history of voice disorders.

### *Analysis*

A neural network (NN) was trained using simulated data from a symmetric triangular body-cover vocal fold model that has a zipper-like closure, posterior glottal opening, three-way interactions between tissue, airflow, and sound, and the independent activation of all five intrinsic laryngeal muscles. The simulated glottal airflow is processed to obtain 10 aerodynamic measures that represent the NN input. The NN architecture consists of two hidden layers with 150 interconnected neurons each. Training was performed with the Levenberg–Marquardt backpropagation algorithm on 80% of the simulations.

## RESULTS

The estimates yielded a mean absolute error (MAE) of 0.12 kPa for 20% synthetic validation data. The performance was also assessed against intraoral pressure in consecutive /pae/ utterances in a case study. Subglottal pressure was estimated using aerodynamic parameters from inverse filtering of oral velocity volume in a 50 ms sliding window, resulting in a MAE of 0.05 kPa.

## DISCUSSION AND CONCLUSION

Our preliminary results illustrate that combining numerical simulations and machine learning tools allows for the non-invasive estimation of subglottal pressure in running speech with a performance similar to prior studies [1]. The approach can be combined with the IBIF algorithm to produce additional features for enhancing the ambulatory assessment of vocal function.

## 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, FONDECYT 1191369 and Beca Doctorado Nacional 21190074.

## REFERENCES

[1] Alzamendi G, Manríquez R, Hadwin P, et al. Bayesian estimation of vocal function measures using laryngeal high-speed videoendoscopy and glottal airflow estimates: An in vivo case study. The Journal of the Acoustical Society of America. 2020; 147, EL434-EL439.

[2] Gomez P, Schutzenberger A, Semmler M, Dollinger M. Laryngeal Pressure Estimation with a Recurrent Neural Network. IEEE J Transl Eng Health Med. 2018; 7:2000111.

[3] Zhaoyan Zhang. Estimation of vocal fold physiology from voice acoustics using machine learning. The Journal of the Acoustical Society of America 2020.; 147, EL264-EL270.

# A NONRANDOMIZED TRIAL FOR STUDENT TEACHERS OF AN IN-PERSON AND TELEPRACTICE GLOBAL VOICE PREVENTION AND THERAPY MODEL WITH ESTILL VOICE TRAINING ASSESSED BY THE VOICEEVALU8 APP

**Elizabeth U. Grillo, Ph.D., CCC-SLP, EFP, CHSE[1]**

[1] Department of Communication Sciences and Disorders, West Chester University, West Chester, PA, USA

***Keywords*: telepractice, voice therapy, mobile voice assessment, app technology**

## INTRODUCTION

This study investigated the effects of the in-person and telepractice (i.e., delivery of speech-language pathology services at a distance via synchronous and asynchronous methods) Global Voice Prevention and Therapy Model (GVPTM) treatment conditions and a control condition with vocally healthy student teachers.[1] The GVPTM includes four components: 1) stimulability testing, 2) bottom-up treatment hierarchy, 3) "new" vs "other/old" voices at each step of the hierarchy, and 4) additional methods that augment and support the target voice productions (e.g., vocal hygiene and education).[1] The GVPTM uses Estill Voice Training (EVT) to guide stimulability, which is a framework that defines auditory-perceptual targets by the anatomy and physiology of the system.[2] For example, oral twang is an auditory-perceptual term defined by the following anatomic and physiologic components; aryepiglottic sphincter narrow, false vocal fold retraction, thyroid tilt, and velum high.

All four components of the GVPTM were applied to the in-person and telepractice treatment conditions. Only the additional methods component of the GVPTM, focusing on vocal hygiene and education, was applied to the control condition. The primary aim of this study was to determine if both the in-person and telepractice GVPTM would demonstrate an increase in the primary outcome measure of fundamental frequency ($f_o$) from pre to post in fall and if that $f_o$ increase would be maintained during student teaching. It was hypothesized that $f_o$ would increase in the treatment conditions because the in-person and telepractice GVPTM included vocal training of a new "resonant" voice in connected speech that eliminated slack (glottal fry), which facilitates an increase in $f_o$.[3,4] In addition, it was hypothesized that the in-person and telepractice treatment conditions would produce similar $f_o$ results in fall and spring. Because this study was the first to assess vocally healthy student teachers across days, weeks, and months in twice daily voice measures using the VoiceEvalU8 app, the secondary outcome measures reflect a comprehensive view of acoustic, perceptual, and aerodynamic measures that will guide continued analysis of the current data and inform future longitudinal research.

## METHODS

In this single-blinded, nonrandomized trial, 82 vocally healthy student teachers completed all aspects of the study. All participants who provided informed consent met the following inclusion criteria; student in a Bachelor of Education program at West Chester University with student teaching planned for spring, owner of either an android or iOS smartphone, and vocally healthy as determined by no current voice complaints and no abnormal voice patterns perceptually judged by the researchers. Participants were selected into either treatment or control conditions based on their availability for weekly training sessions. The in-person condition met weekly for 45-60 minutes across four weeks with week 1 covering vocal hygiene and education and weeks 2-4 providing vocal training. For telepractice, the online learning management system called Desire 2 Learn (D2L) included all synchronous and asynchronous methods for weeks 1-4. For week 1, participants completed vocal hygiene and education via asynchronous content and assignments. During weeks 2-4, vocal training occurred weekly for 45-60 minutes via synchronous videoconferencing. For the control condition, participants completed only vocal hygiene and education asynchronously though D2L.

### *Outcome Measures*

$f_o$ was the primary acoustic outcome measure and was captured twice daily in three 5 sec. trials of sustained /a/, the phrase "we were away a year ago," and a 15-sec. connected speech sample. The secondary acoustic outcome measures were jitter%, shimmer %, noise-to-harmonic ratio (NHR), cepstral peak prominence, smoothed cepstral peak prominence, and acoustic voice quality index captured twice daily in the same utterances used for the $f_o$ analysis. The secondary perceptual outcome measures were the Voice Handicap Index (VHI)-10 and Vocal Fatigue Index (VFI) completed through the app once a week rather than twice a day so that participants were not desensitized to the scales. The secondary aerodynamic outcome measures were the s/z ratio and maximum phonation time (MPT) captured twice a day on days 1, 3, and 5 for the s/z ratio and on days 2 and 4 for MPT.

*Analysis*

The dependent variables were the primary and secondary outcome measures captured by VoiceEvalU8. Independent variables (IV) were year (i.e., year 1, 2, 3, and 4), condition (i.e., treatment condition and control condition), time (i.e., data collection time points for fall and spring semesters), twice daily logs (i.e., 6-11 a.m. and 4-11 p.m.), and day (i.e., Monday-Friday). The current analysis focused on two of the IVs, condition, and time; therefore, two-way analysis of variances (ANOVA) was used with an alpha level of 0.05. For post-hoc testing, the Bonferroni method was applied only for significant main and interaction effects adjusting alpha level for multiple pairwise comparisons. For fall, the focus of the current analysis was the condition by time significant interaction effects with post hoc testing to determine significant differences between conditions at post. For spring, there were nine weeks for the time IV with no pre and post; therefore, the focus of the current analysis was the significant main effects of condition only with post hoc testing to determine if the three conditions overall had a different effect on the outcome measures.

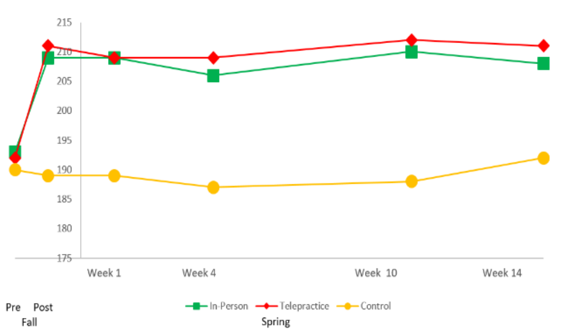



Figure 1. Mean $f_o$ (Hz) in the phrase "we were away a year ago" for females at fall pre and post and spring while student teaching for 14 weeks across in-person, telepractice, and control conditions.

## Results and Discussion

The in-person and telepractice GVPTM facilitated a new voice for connected speech that was more resonant by eliminating slack true vocal fold body-cover (glottal fry), retracting false vocal folds, tilting thyroid, and adding a head/neck anchor, if needed. The improvement in voicing was captured by VoiceEvalU8 in the treatment conditions through $f_o$ in fall from pre to post and maintained in spring while the participants were student teaching. The in-person and telepractice conditions produced the same increase in mean $f_o$ indicating that both treatments were effective. The control condition did not demonstrate improvements in voicing as evidenced by a stable $f_o$ in fall and during spring because participants in the control condition did not receive vocal training. Overall, the secondary acoustic outcome measures during sustained /a/ did not capture any voice changes. The phrase and speech utterances were more successful in showing voice changes for some of the secondary acoustic outcomes measures (i.e., NHR in the phrase and speech and jitter% in speech) across conditions in fall. The perceptual measures of the VHI-10 and VFI did not document changes in voice related quality-of-life (QOL) and vocal fatigue at fall pre and post. However, during the spring while the participants were student teaching, the VHI-10 and the VFI were successful in documenting an increased negative impact on voice related QOL and vocal fatigue in the control condition as compared to the treatment conditions suggesting that clinicians and researchers need to assess voice beyond the typical "snapshots" at just pre and post to a "landscape" view across weeks and months. The aerodynamic measures of s/z ratio and MPT were not successful in documenting voice changes in fall and spring possibly due to vocal training focusing on new voices for connected speech, falsetto, oral twang, and belt rather than on sustaining phonemes. Outcome measures should reflect the goals/targets of the prevention/therapy model.

In sum, the in-person and telepractice GVTPM treatment conditions were successful in improving the participants' voices across acoustic measures and maintaining that improvement during student teaching. In addition, the in-person and telepractice GVPTM treatment conditions were successful in decreasing the negative impact of voice related QOL and vocal fatigue during student teaching.

## Acknowledgments

The work was supported by the National Institute on Deafness and Other Communication Disorders, R15DC014566. The author would like to thank the graduate student clinicians who conducted vocal training with the participants. Thank you to Dr. Robert Gallop who assisted with data analysis and to Dr. Jarrad Van Stan who provided feedback on applying the GVPTM to the Rehabilitation Treatment Specification System.

## References

[1] Grillo, E.U. (2021). A nonrandomized trial for student teachers of an in-person and telepractice Global Voice Prevention and Therapy Model with Estill Voice Training assessed by the VoiceEvalU8 app. *American Journal of Speech-Language Pathology*, Feb 1, 1-18.

[2] Steinhauer, K., Klimek, M. M., & Estill, J. (2017). *The Estill voice model: Theory & translation*. Pittsburgh, PA: Estill Voice International.

[3] Barone, N.A., Ludlow, C.L., & Tellis, C.M. (2019). Acoustic and aerodynamic comparisons of voice qualities after vocal training. *Journal of Voice*; Sep 3; S0892-1997(19)30033-5.

[4] Plexico L.W. & Sandage M.J. (2017). Influence of glottal fry on acoustic voice assessment: A preliminary study. *Journal of Voice*; 31(3): 378.e13-378.e17.

# ACCURACY OF ACOUSTIC MEASURES OF VOICE VIA TELEPRACTICE VIDEOCONFERENCING PLATFORMS

**Hasini R. Weerathunge[1, 2], Roxanne K. Segina[2], Lauren Tracy[3], Cara E. Stepp[1, 2, 3]**

[1] Department of Biomedical Engineering, Boston University, Boston, MA
[2] Department of Speech, Language, and Hearing Sciences, Boston University, Boston, MA
[3] Department of Otolaryngology-Head and Neck Surgery, Boston University School of Medicine, Boston, MA

***Keywords*: Acoustics; Dysphonia; Voice Disorders; Telepractice**

## INTRODUCTION

Telepractice improves patient access to clinical care for voice disorders. Acoustic assessment has the potential to provide critical, objective information during telepractice, yet its validity via telepractice is currently unknown. The current study investigated the accuracy of acoustic measures of voice in a variety of telepractice platforms. All acoustic measures explicitly based on noise, or based on signal perturbation, were hypothesized to be significantly impacted by transmission condition with large effect sizes, whereas time-based measures were hypothesized to be impacted but with a small effect size, due to the lack of explicit reliance on noise.

## METHODS

### *Participants*

Voice samples from a group of 29 cisgender participants (female = 14, male = 15) with a variety of voice disorder diagnoses and over a large age range (19 – 82 years) were selected from an existing database of over 1,400 participant speech samples. All participants were speakers of American English with no other history of speech, language, or hearing disorders.

### *Instrumentation and Measurement*

Voice samples were transmitted over six video conferencing platforms (Zoom with and without enhancements, Cisco WebEx, Microsoft Teams, Doxy.me, and VSee Messenger). Each platform was chosen based on their adherence to HIPPA standards and their prevalence of use for voice telepractice [1]

### *Analysis*

Standard time-, spectral-, and cepstral-based acoustic measures were calculated. The effect of transmission condition on each acoustic measure was assessed using repeated measures analyses of variance. For those acoustic measures for which transmission condition was a significant factor, linear regression analysis was performed on the difference between the original recording and each telepractice platform, with the overall severity of dysphonia, internet speed, and ambient noise from the transmitter as predictors.

## RESULTS

Transmission condition was a statistically significant factor for all acoustic measures except for mean vocal fundamental frequency ($f_o$). Ambient noise from the transmitter was a significant predictor of differences between platforms and the original recordings for all acoustic measures except $f_o$ measures. All telepractice platforms affected acoustic measures in a statistically significantly manner, although the effects of platforms varied by measure.

## DISCUSSION AND CONCLUSION

Overall, measures of $f_o$ (mean, standard deviation, range) were the least impacted by telepractice transmission. Sound pressure level variability and acoustic measures aimed at voice quality were impacted by most telepractice platforms. Changes in acoustic measures of voice quality due to transmission were as large or larger than differences reported between individuals with and without voice disorders in previous work, suggesting that telepractice platform transmission imposes clinically relevant degradations to these measures. Microsoft Teams had the least and Zoom used with enhancements had the most pronounced effects on acoustic measures overall. These results provide valuable insight into the relative validity of acoustic measures of voice when collected via telepractice.

## ACKNOWLEDGMENTS

Research funded by research grant R01 DC015570 and UL1 TR001430 from the National Institutes of Health. Special thanks to D. Buckley and K. Dahl for completing auditory-perceptual ratings of stimuli.

## REFERENCES

[1] United States, Department of Labor, Employee Benefits Security Administration, *Health Coverage Portability: Health Insurance Portability and Accountability Act of 1996 (HIPAA)*. US Department of Labor, Employee Benefits Security Administration, 2004.

*The 14th International Conference on Advances in Quantitative Laryngology, Voice and Speech Research (AQL)*

# PSYCHOMETRIC ANALYSIS OF AN ECOLOGICAL VOCAL EFFORT SCALE IN INDIVIDUALS WITH AND WITHOUT VOCAL HYPERFUNCTION DURING ACTIVITIES OF DAILY LIVING

**Katherine L. Marks[1,2,3*], Alessandra Verdi[1,2], Laura E. Toles[1,2], Kaila L. Stipancic[1,5], Andrew J. Ortiz[2,4], Robert E. Hillman[1,2,4], Daryush D. Mehta[1,2,4]**

[1] MGH Institute of Health Professions, Boston, MA, USA
[2] Massachusetts General Hospital, Boston, MA, USA
[3] Boston University, Boston, MA, USA
[4] Harvard Medical School, Boston, MA, USA
[5] University at Buffalo, Buffalo, NY, USA

***Keywords*: Voice; Vocal Effort; Psychometrics; Ambulatory Voice Monitoring**

## INTRODUCTION

One of the most frequent complaints of patients with vocal hyperfunction is the requirement of increased vocal effort to speak [1], which is defined as the perception of the work or exertion an individual feels during phonation [2]. The aim of the current study was to examine the ecological momentary assessment of vocal effort in an individual's real-world speaking environment. We examined the psychometric properties (reliability, validity, sensitivity, and responsiveness) of an ecological vocal effort scale that was temporally linked to a voicing task and used to capture vocal effort ratings throughout a week of ambulatory voice monitoring in individuals with and without vocal hyperfunction.

## METHODS

Thirty-eight patients with phonotraumatic vocal hyperfunction (PVH), 17 patients with non-phonotraumatic vocal hyperfunction (NPVH), and 39 vocally healthy controls participated in a week of smartphone-based ambulatory voice monitoring. Following the methodology of Van Stan and colleagues [3], participants were prompted throughout the day to answer a global vocal status question, produce a consonant-vowel syllable string, and rate the vocal effort needed to produce the task on a visual analog scale (0–100). Each hour, they were asked whether their global vocal status had improved, worsened, or not changed. If they reported a change, they were asked to repeat the syllable string and re-rate the effort needed to produce the task.

## RESULTS AND DISCUSSION

Reliability: The overall intraclass correlation coefficient, ICC(A,1) was 0.96, indicating excellent test-retest reliability. The standard error of measurement (SEM) was found to be 4.14. Validity: Welch's F revealed a statistically significant main effect of diagnosis on week-long mean vocal effort scores ($F(2,19) = 13.44$, $p < .001$, $\eta^2 = .59$). Post hoc pairwise comparisons revealed large effect sizes between the PVH and controls ($p < .01$, $d = 1.62$) and between the NPVH group and controls ($p < .01$, $d = 1.61$). *Sensitivity to Change:* The ecological vocal effort scale was highly sensitive in discriminating individuals with PVH and NPVH from controls ($d = 1.62$) and patients pre-treatment to post-treatment ($d = 1.75$). The SEM was used to obtain a minimal detectable change value of 12 scalar points. *Responsiveness:* The minimal clinically important difference was 9.7, which is within the error of the measure. Therefore, we must rely on the minimal detectable change as a threshold for real change.

## CONCLUSION

In the context of ambulatory voice monitoring, the ecological vocal effort scale linked to a voicing task was found to be reliable, valid, and sensitive to the presence of vocal hyperfunction and to successful voice treatment in individuals with vocal hyperfunction. Future work may determine whether the changes in vocal effort are related to vocal behaviors by investigating the objective ambulatory voice measures.

## ACKNOWLEDGMENTS

Funding provided by the Voice Health Institute and the National Institutes of Health (NIH) NIDCD (P50 DC015446). The paper's contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH. The authors thank Drs. Jarrad Van Stan, Annie Fox, Alan Jette, James Burns, and Tiffiny Hron, and Ms. Brianna Williams and Ms. Liane Houde for contributions. Disclosure: Drs. Robert Hillman and Daryush Mehta have a financial interest in InnoVoyce LLC, a company focused on developing and commercializing technologies for the prevention, diagnosis, and treatment of voice-related disorders.

## REFERENCES

[1] Van Mersbergen *et al. Perspectives of the ASHA Special Interest Groups* 2020; 6(1): 69-79.
[2] Hunter *et al. Journal of Speech, Language, and Hearing Research* 2020; 63(2): 509-532.
[3] Van Stan *et al. American Journal of Speech-Language Pathology* 2017; 26(4): 1167-1177.

# POSTURE AND PHONATION: SUPRAHYOID, INFRAHYOID, AND RESPIRATORY MUSCLE ACTIVITY DURING BODY POSITION MODIFICATIONS.

**Adrián Castillo-Allendes[1], Mauricio Delgado-Bravo[2], Alvaro Reyes[3], Eric J. Hunter[1]**

[1] Department of Communicative Sciences and Disorders, Michigan State University, East Lansing, Michigan, USA
[2] Departamento de Ciencias de la Salud, Facultad de Medicina, Pontificia Universidad Católica de Chile, Santiago, Chile
[3] Faculty of Rehabilitation Sciences, Universidad Andres Bello, Viña del Mar, Chile

***Keywords*: surface EMG, voice therapy, body posture, respiratory muscles**

## INTRODUCTION

Respiratory and extrinsic laryngeal muscles are essential for phonation contributing to subglottal pressure generation and laryngeal stability. Respiratory and extrinsic laryngeal muscles could be affected by postural changes which have been used by voice teachers and clinicians as a strategy to promote healthy and economical voice production [1,2]; however, postural changes have been inconsistently described in the literature, leaving room for free interpretation and possible misunderstandings as a result of lack of biomechanical foundations to support this approach [3]. There is no previous description of phonatory muscle activity using surface electromyography during voice production in different body posture in healthy volunteers. Knowing which postures generate greater activation of the respiratory muscles could be beneficial to justify using these strategies as a tool for the voice approach.

The purpose of this study was to compare the activation magnitude of the muscles involved in the phonation-breathing function using surface electromyography (sEMG) in four body postures.

## Methods

### *Participants*

Eight healthy voice speakers (age range 19-35 years) with at least three years of vocal training and no musculoskeletal pathologies were recruited in this study.

### *Instrumentation and Measurement*

Surface electromyography (sEMG) activity was captured using a 16 channels Surface EMG system (Delsys Trigno Wireless System, Boston, USA). Muscles analyzed were suprahyoid (SH), infrahyoid (IH), scalene (S), sternocleidomastoid (SCM), upper trapezium (UT), rectus abdominis (RA) and lumbar multifidus (LM). Maximum voluntary isometric contraction (MVIC) of all muscles was recorded and used for signal normalization. The percentage of muscle activity was calculated and recorded for analysis.

Each participant performed 2 phonatory tasks (the production of an /a:/ as a staccato (five times), and an /a:/ sustained for 10 seconds) in four different postures: upright (P1), modified upright (P2), leaning “tower of Pisa” like posture (P3), and upright and standing on an unstable surface (P4). The phonatory tasks were controlled using a tempo of 30 BPM, between 80-100 dB, and at note A3 (220 Hz) and A4 (440 Hz) by males and females, respectively.

### *sEMG signal Analysis*

For each sEMG signal, a 20 Hz digital fourth-order Butterworth high-pass filter was performed to suppress the postural and voice noise from motion artifacts. Then, a digital bandstop filter (220 Hz for males, 440 Hz for females) was performed to suppress the noise effect from each phonatory task over EMG signals. Finally, an RMS calculate was performed for signals of the two phonatory tasks, with an analysis window of 500 ms. The sEMG activity was expressed as percentage of MVIC.

### *Statistical Analysis*

One way analysis of variance ANOVA was used to compare sEMG activity of phonatory muscles between postures. Linear regression analysis was used for multiple comparisons analysis and upright posture (P1) was used as baseline. Statistical significance was set at $p<0.05$.

## Results

In the production of staccato, UT showed a significant reduction ($p=0.049$) in the percentage of sEMG activity in the modified upright position (mean:2.32; SD:2.04) compared to the upright position (mean:3.29; SD: 1.96). RA muscles showed a significant reduction in the percentage of sEMG activity in the modified upright (mean: 3.80; SD: 3.58; $p=0.021$) and leaning position (mean: 3.88; SD: 3.16; $p=0.021$) compared to the upright position (mean:5.30; SD: 3.31). No significant differences were detected between postures in the performance of the sustained vowel /a:/ in the percentage of sEMG activity.

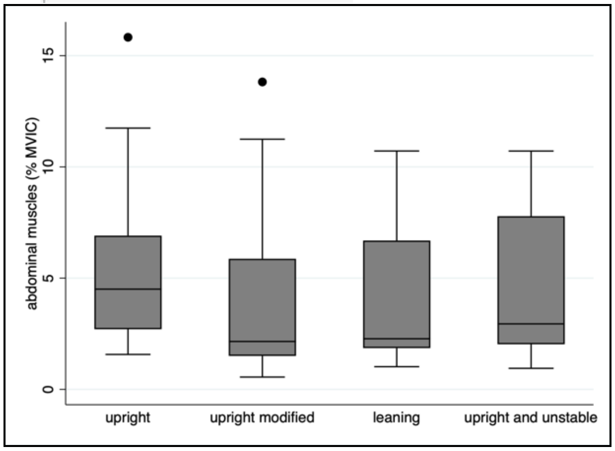



Figure 1. Percentage of sEMG activation of the abdominal muscle. The X-axis represents each of the postures tested.

### DISCUSSION

Postural modifications have been used to obtain better results in voice production due to a possible increase in the respiratory musculature activity, increased management of the respiratory support (appoggio), and a decrease in the neck muscles activity [4]. Regarding sEMG activity, although we have observed some effects on the abdominal and upper trapezius muscles, it was not possible to identify whether one particular posture was more beneficial than another for such effects. Future studies should consider a larger number of subjects, respiratory dynamics, and voice quality. Concerning muscle activity, we have observed some effects on the RA and UT muscles; it was not possible to identify whether one particular posture was more beneficial increasing the muscle activity than another for such effects. Future studies should consider a larger number of subjects, respiratory dynamics, and voice quality.

### CONCLUSION

The postural modifications in this study did not show an increase in sEMG activity involved in phonation-breathing tasks that would justify the use of these strategies as a tool for voice approach. UT and RA decreased their muscle activity in P2 posture.

### REFERENCES

[1] Wilson Arboleda BM, Frederick AL. Considerations for Maintenance of Postural Alignment for Voice Production. *J Voice*. 2008 Jan;22(1):90–9.

[2] Doscher B. *The Functional Unity of the Singing Voice*. Vol. 1. Scarecrow Press; 1993 [cited 2021 Feb 1]. 352 p.

[3] Emerich Gordon K, Reed O. The Role of the Pelvic Floor in Respiration: A Multidisciplinary Literature Review. *J Voice*. 2020 Mar;34(2):243–9.

[4] Hamdan AL, Deeb R, Tohme RA, Rifai H, Husseini S, Fuleihan N. Vocal technique in a group of Middle Eastern singers. *Folia Phoniatr Logop*. 2008;60(4):217–21.

# ASYMMETRIC SUPERIOR LARYNGEAL NERVE STIMULATION: INFLUENCE ON ACOUSTIC CEPSTRAL PEAK PROMINENCE AND GLOTTIS SYMMETRY AND PERTURBATION MEASURES

**Patrick Schlegel[1], Dinesh Chhetri[1]**

[1] Department of Head & Neck Surgery, UCLA, Los Angeles, California, United States of America

***Keywords*: Voice; Laryngeal Nerve Stimulation; High-Speed Videoendoscopy; Acoustics**

## INTRODUCTION

During healthy phonation the vocal folds, producing the acoustic source signal, are expected to oscillate steadily and symmetrically. An asymmetric vibration pattern is often seen in laryngeal voice disorders [1]. However, how exactly asymmetric vibration influences the acoustic signal is still not well understood.

One approach to investigate coherencies between specific vocal fold oscillation patterns and the acoustic signal is artificial stimulation of the superior and recurrent laryngeal nerves (SLN and RLN) [2]. By using parallel acoustic and high-speed video recording, coherencies between different features of the oscillation process can be investigated. From the video data the "Glottal area waveform" (GAW) can be extracted, reflecting the change of the area between the vocal folds over time. Afterwards certain features as periodicity or symmetry can be determined from total or partial GAWs and linked to features of the acoustic signal.

## METHODS

### *In vivo canine model*

For this work raw data collected in previous work was used. A comprehensive description of the canine model and recording process can therefore be found in [2] and only a short summary is given here: Surgical exposure of laryngeal nerves in a mongrel canine was prepared. Left and right RLNs and SLNs were stimulated. RLNs were stimulated to achieve complete and symmetric closure. RLN activation levels remained the same for all conditions. SLNs were stimulated separately over eight equidistant levels of activation, ranging from 0 (no activation) to 7 (maximum activation). This resulted in 64 combinations of SLN activation with constant RLN.

Phonation was achieved by providing rostral airflow using a subglottal tube attached to the trachea. Airflow was increased linearly from 300 to 1,400 milliliters/second (ml/s) over a period of 1.5 seconds during nerve stimulation [2].

### *Recording process*

Vocal folds oscillation was recorded using high-speed video recording at 3000 frames per second (fps). The acoustic signal was recorded using a probe tube microphone at 50,000 fps. The microphone was mounted with the inner wall of the subglottic inflow tube below the glottis [2].

### *Segmentation*

The glottal area between the vocal folds was segmented using a custom version of the software "Glottis Analysis Tools 2020" (GAT) featuring a static midline. For a detailed description of the segmentation process see [3]. As can be seen in Fig. 1 the constantly open posterior section of the glottis was excluded during segmentation, as for this study we were only interested in the changing area. A static midline was drawn for each video. For a 1000 frame section (flow = ~1,050-1,300 ml/s) of each video total, left and right GAWs were extracted.

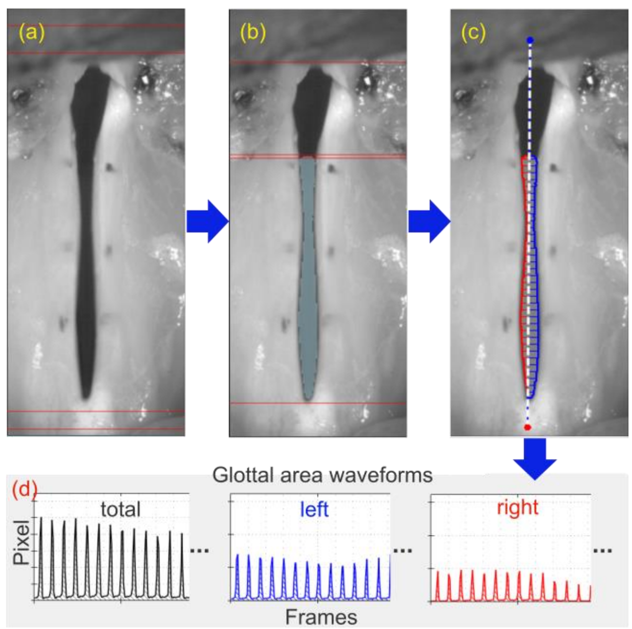



Figure 1. (a) select region of interest (b) segment area (c) draw midline (d) extract GAWs based on 1000 frames.

### *Parameter calculation*

Maximum based Cycles were detected using MATLAB (version R2020b) for peak detection. Using GAT, Jitter % and Shimmer % were calculated on the total GAW, Phase Asymmetry was calculated using the partial GAWs. Using Praat (version 6.1.08), Smoothed Cepstral Peak Prominence (CPPS) was calculated on the acoustic signal.

## Results

Phase Asymmetry reflected asymmetry in SLN stimulation (Fig 2). Jitter % and Shimmer % increased and decreased over time resulting in strongly increased values for SLN-stimulation levels between 3 and 6 (see Fig 3). Acoustic CPPS correlated with Jitter % (-0.643, p<10^-7) and Shimmer % (-0.641, p<10^-7) but not with Phase Asymmetry (-0.025, p>0.05) (Fig 4).

## Discussion

Phase asymmetry reflects the asymmetric SLN activation and GAT measurements of asymmetry are also in agreement with manually derived asymmetry ratings reported in an earlier experiment using the same raw data [2]. There are still potential influencing factors on this parameter e.g., small movements of the entire glottis, but it may be a valuable tool to objectively grade oscillation asymmetry in similar settings.

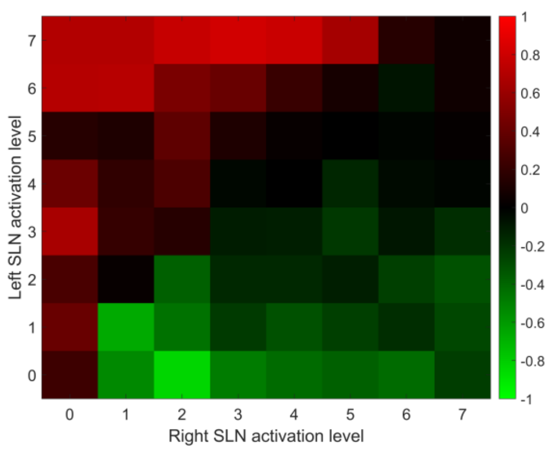



Figure 2. Phase Asymmetry for all 64 combinations of left and right SLN stimulation.

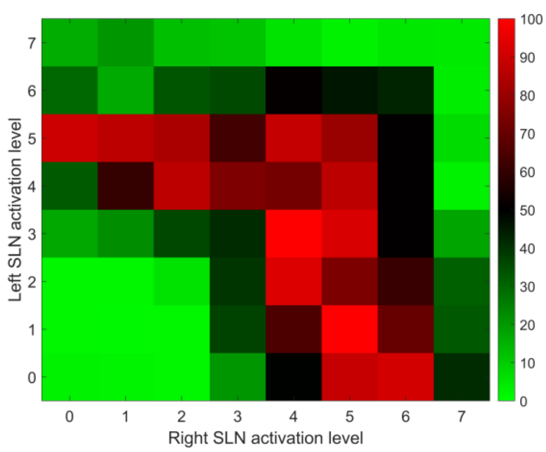



Figure 3. GAW Jitter % for all 64 combinations of left and right SLN stimulation.

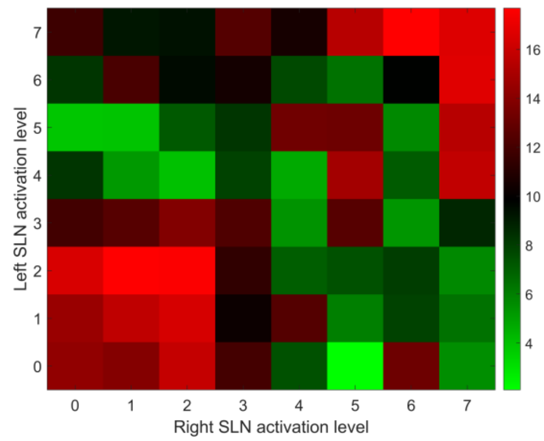



Figure 4. Acoustic CPPS % for all 64 combinations of left and right SLN stimulation.

The increase of Jitter and Shimmer may be related to a complex interaction of RLN and SLN activation levels leading to overall more or less stable phonation conditions. However, since the increase and decrease of perturbation happened for consecutive recordings it may also be related to other factors that were not directly related to RLN or SLN stimulation.

CPPS reflects the increase of Jitter and Shimmer but does not reflect asymmetry, as the negligible correlation between Phase Asymmetry and CPPS as well as the comparison of Fig 2 and Fig 4 shows. However, it may be that Jitter and Shimmer mask potential small coherencies.

## Conclusion

More acoustic measures must be investigated to explore how vocal fold oscillation asymmetry affects the acoustic signal, preferably measures robust to high Jitter and Shimmer.

## Acknowledgments

This project was funded by Grant No R01 DC11300 from the NIH/NIDCD.

## References

[1] Inwald E, Döllinger M, Schuster M, Eysholdt U, Bohr C. Multiparametric analysis of vocal fold vibrations in healthy and disordered voices in high-speed imaging. Journal of Voice. 2011; 25(5):576–590

[2] Chhetri DK, Neubauer J, Bergeron JL, Sofer E, Peng KA, Jamal N. Effects of Asymmetric Superior Laryngeal Nerve Stimulation on Glottic Posture, Acoustics, Vibration. Laryngoscope 2013; 123:3110–3116.

[3] Schlegel P, Kniesburges S, Dürr S, Schützenberger A, Döllinger M. Machine learning based identification of relevant parameters for functional voice disorders derived from endoscopic high-speed recordings. Scientific Reports. 2020; 10:10517.

# EFFECT ON VOCAL DOSES OF A WORKPLACE VOCAL HEALTH PROMOTION PROGRAM

**Ángela Patricia Atará-Piraquive[1], Lady Catherine Cantor-Cutiva [1]**

[1] Department of Collective Health, Universidad Nacional de Colombia, Bogotá, Colombia

***Keywords*: Occupational Voice Users; Vocal Doses; College Professors; Vocal Health**

## INTRODUCTION

College professors could be considered as occupational voice users who have high voice use with moderate vocal quality (1). College professors are exposed to many environmental and organizational factors in the workplace that might affect their vocal demand response. Previous studies have reported that college professors work in noisy environments with large numbers of students (2,3). They may work in more than one university (4) and carry out research and administrative activities that could increase vocal demand (5). For this reason, it is necessary to carry out actions focused on helping college professors to reduce vocal demand in class settings. Likewise, highlight the importance of promoting healthy vocal habits within the workplace by implementing programs to promote vocal health and improve vocal production in classrooms.

Traditionally, vocal health programs usually have been focused on vocal hygiene (6), vocal training with resonant exercises (7), instructions about how to adopt good posture (8) and no included teaching-learning or organizational strategies about vocal health in workplaces. Therefore, Workplace Vocal Health Promotion Programs are a good option to take actions to promote healthy vocal habits for college professors. Currently, there are programs with tele-practice and in-person services (9). In this way, the objective of this study was to analyse changes in vocal doses due to the implementation Workplace Vocal Health Promotion Program for college professors.

## METHODS

*Participants*

The participants were randomly assigned into two groups: Intervention group (n=6) or No intervention group (n=4). They were college professors from public university in Colombia, who met the following inclusion criteria 1) being professor in the public university, 2) have maximum 4 hour of frontal lectures per day, and 3) have not received any voice therapy or training during the last three years.

*Instrumentation and Measurement*

Each participant filled out a general questionnaire (sociodemographic information, vocal habits, the presence of medical conditions and extra-work activities related with intensive voice use).

Before and after the intervention, the participants were monitored in one of their lectures using a vocal dosimeter through vocal dosimeter build up with a collar microphone and TASCAM DR-05 recorder; and filled in self-report questionnaire of vocal symptoms.

The program contained 2 face-to-face sessions and 2 virtual sessions, which were developed using a platform called "Moodle". Each session was performed one per week with 45 minutes of duration. Four modules were proposed in this intervention: (1) vocal hygiene, (2) vocal training, and (3) organizational strategies to be implemented in the workplace for Intervention Group (Figure 1). No intervention group received the virtual sessions after completing the follow-up.

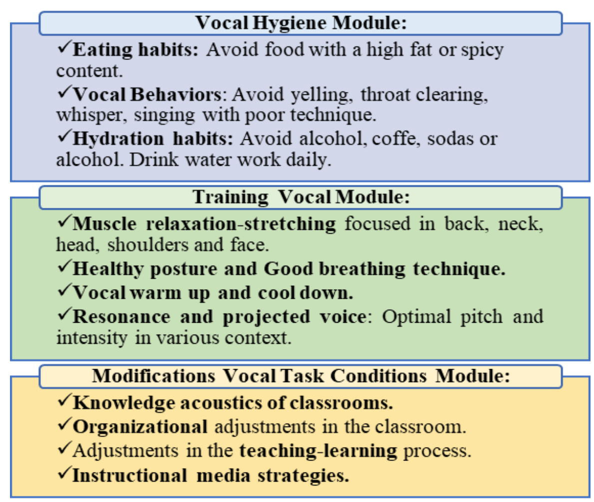



Figure 1. Description modules of the Workplace Vocal Health Promotion Program

*Analysis*

The Shapiro Wilk test was used to determine the normality of the data prior to statistical analysis. We used the Wilcoxon signed-rank test to assess the difference in vocal doses between before and after intervention. The Mann-Whitney U Test was used to determine the difference in vocal doses between the Intervention and No-Intervention groups.

## RESULTS

There were four males and two females in the Intervention group, while there were two males and two females in the No intervention group. Participants in the Intervention group averaged 35.3 years old (SD= 6.3), while subjects in the No-intervention group averaged 49.2 years old (SD=16.1). There was a tendency of decreasing time dose of around 3% among participants in the intervention group, while the professors from the No-intervention group showed an increase of 1% (p >0.05; Wilcoxon signed-rank test). For the intervention group, there was a reduction in the VLI and Dd_n; while for the No intervention group there was an increase in Dd_n at the end of the study but without statistically significant differences (p> 0.05; Wilcoxon signed-rank test).

**Table 8.Vocal doses before and after intervention**

| | Intervention Group (n=6) | | | No intervention Group (n=4) | | |
|---|---|---|---|---|---|---|
| | PRE Media | POST Media | *p-value* | PRE Media | POST Media | *p-value* |
| Dt (%) | 37,3 | 34,8 | 0,06 | 47,9 | 49,1 | 0,23 |
| VLI | 402,7 | 362,7 | 0,13 | 592,7 | 543,2 | 0,14 |
| Dd_n | 4,15 | 4,10 | 0,87 | 3,75 | 3,86 | 0,68 |

Dt: Time dose; VLI: Vocal Loading Index; Dd_n: Normalized Distance dose. Wilcoxon signed-rank test. $p < 0.05$

## DISCUSSION AND CONCLUSION

College professors are occupational voice users exposed to different risk factors that can directly increase their response to voice demand. Before intervention program, participants showed higher time dose values (35-49%) than public school teachers, who had 33% time dose (10). This may mean that college professors have a high vocal load represented in the vocal demands that demand to speak in a stage with acoustic challenges, high numbers of students and have two-hour lectures.

The Workplace Vocal Health Promotion Program showed a positive effect on vocal demand represented in reduction of time dose, VLI and Dd_n, which means that there is a decrease in the amount of vocal folds vibration. Therefore, a reduction in the amount of vibration of the vocal cords is inferred, and a decrease in damage to the vocal cords because of the collision force based on the findings of previous studies (11,12). Even though there were no significant changes in the vocal doses with the implementation of the program, there can be positive effects in the reduction of the vocal load. Interventions in the field of vocal health must include strategies aimed at the organizational component as well as applied in a module of the program for teachers. Virtual sessions can be a good option for teachers to adopt healthy vocal habits, if they have the accompaniment of the clinician.


## ACKNOWLEDGMENTS

Special thanks to Professor Ximena Saenz Montoya for providing us with the Laboratory of The Nursing Faculty to carry out the assembly of the vocal dosimetry equipment.


## REFERENCES


[1] Vilkman E. Voice problems at work: A challenge for occupational safety and health arrangement. Folia Phoniatr Logop. 2000;52(1-3):120-5.

[2] Korn GP, Augusto de Lima Pontes A, Abranches D, Augusto de Lima Pontes P. Hoarseness and Risk Factors in University Teachers. J Voice. 2015;29(4): 518.e21-28.

[3] Gomes NR, Teixeira LC, de Medeiros AM. Vocal Symptoms in University Professors: Their Association with Vocal Resources and with Work Environment. J Voice. 2020;34(3):352-7.

[4] Cantor-Cutiva LC, Robles-Vega HY, Sánchez EA, Morales DA. Differences on Voice Acoustic Parameters between Colombian College Professors with and without Vocal Fatigue. J Voice. 2020.

[5] Kyriakou K, Petinou K, Phinikettos I. Risk Factors for Voice Disorders in University Professors in Cyprus. J Voice. 2018;32(5): 643.e1-643.e9.

[6] Boominathan P, Chandrasekhar D, Nagarajan R, Madraswala Z, Rajan A. Vocal Hygiene Awareness Program for Professional Voice Users (Teachers): An Evaluative Study from Chennai. Asia Pac J Speech Lang Hear. 2008;11(1):39-45.

[7] Roy N, Gray SD, Simon M, Dove H, Corbin-Lewis K, Stemple JC. An evaluation of the effects of two treatment approaches for teachers with voice disorders: a prospective randomized clinical trial. J Speech Lang Hear Res. 2001;44(2):286-96.

[8] Rantala L, Sala E, Kankare E. Teachers' Working Postures and Their Effects on the Voice. Folia Phoniatr Logop. 2018;70(1):24-36.

[9] Grillo EU. An Online Telepractice Model for the Prevention of Voice Disorders in Vocally Healthy Student Teachers Evaluated by a Smartphone Application. Perspect ASHA Spec Interest Groups. 2017;2(3):63-78.

[10] Titze IR, Hunter EJ. Comparison of Vocal Vibration-Dose Measures for Potential-Damage Risk Criteria. J Speech Lang Hear Res.2015;58(5):1425.

[11] Verdolini K, Druker DG, Palmer PM, Samawi H. Laryngeal adduction in resonant voice. J Voice. 1998;12(3):315-27.

[12] Assad JP, Magalhães M de C, Santos JN, Gama ACC. Vocal dose: an integrative literature review. Rev CEFAC. 2017;19(3):429-38.

# CONCENTRATION OF HYDROXIDE PEROXIDE IN EXHALED BREATH CONDENSATE AFTER PHONOTRAUMA: A NONINVASIVE TECHNIQUE FOR MEASURE VOCAL INFLAMMATION

**C. Castro[1-2-3], M. Guzman[1], K. Acevedo[3], C. Moran[1], O. Araneda[1], C. Pacheco[4], C. Quezada[2]**

[1] Department of Communication Sciences and Disorders, Universidad de los Andes, Santiago, Chile
[2] Department of Communication Sciences and Disorders, Universidad de Chile, Santiago, Chile
[2], Department of Communication Sciences and Disorder, Universidad de Valparaíso, Valparaíso, Chile
[3] Department of Communication Sciences and Disorder, Universidad San Sebastian, Santiago, Chile
[4] Department of Communication Sciences and Disorder, Universidad de Chile, Santiago, Chile
[5 Clínica] las Condes. Santiago, Chile

***Keywords*: Inflammation, Biomarkers, Vocal loading, Phonotrauma**

## INTRODUCTION

The extensive phonation time causes continuous biomechanical stress on vocal folds. When increases the contact between vocal folds, also increases the structural damage of the mucosa of the vocal folds. This is also called phonotrauma [1],2]. biomarkers such as peroxide (H202) were useful for identifying the vocal folds inflammation process after 4 hours, 8 hours, and 24 hours of vocal fold injury [3]. However, the recollected biomarkers using endoscopy are highly invasive.

Breath condensate (EBC) is a non-invasive technique used to monitor inflammation and oxidative damage in the airway [4]. The system was successful in identifying an inflammatory process in response to intense exercises and hypoxia [5],[6]. We propose extending these efforts by evaluating biomarkers of voice inflammation after vocal loading tasks using the EBC system.

## METHODS

*Participants*

Twenty-nine vocally healthy participants were included and randomized to an experimental and a control group. All participants did not present voice disorders, not smoking, and did not present a history of respiratory disorders. Individuals of the experimental group underwent a 1-hour vocal loading procedure consisting of a loud reading for 15 minutes with 3 minutes of rest. This sequence was repeated four times for a total aloud reading of 60 minutes, the level of reading was 85-90 – dB. This target will be monitored by the same experimenter using a sound level meter positioned 25 cm from the participant's mouth.

*Instrumentation and Measurement*

At baseline, immediately after loading, after the 4-hour, and 24 hours post-baseline, 1.5 ml of EBC was obtained. All samples were stored at -80 °C. Samples were also obtained from control subjects (non-vocal loading) at the same four-time points as the experimental group but without vocal loading. For all samples, the concentration of H2O2 was estimated using a spectrophotometric method based on the oxidation of Iron in an acid medium (FOX reagent).

*Analysis*

Data were fitted by means of the Mixed-Effects Model with Time measures nested within subjects. Tuckey Post-hoc test was performed to study the variation among the factors.

*Results*

A significant interaction was found for the Pre/Post 4h contrast when comparing the Control group and Experimental group ($p=0.049$). A Tukey post-hoc contrast grouping means across measures showed only one homogeneous group for the control condition and two groups for the experimental condition: one including Pre, Post, and Post 24Hrs and another one consisting only of Post 4Hrs.

## DISCUSSION

Preliminary results suggest that the use of EBC is successful for measuring the concentration of h202 linked with inflammation after phonotrauma induced by vocal loading task. The results are consistent with previous studies using biomarkers of vocal folds inflammation, with a maximum peak of concentration at 4 hours after the vocal loading task and a decrease 24 hours after voice rest [2], [3].

## CONCLUSION

Phonotrauma causes an increase in the concentration of H2O2 obtained from EBC at four hours, which is compatible with the generation of an inflammatory process in the vocal folds.

## ACKNOWLEDGMENTS

This work was supported by AINID grant, FONDECYT 11180291

**REFERENCES**

[1] Verdolini K, Rosen CA, Branski RC, Hebda PA. Shifts in biochemical markers associated with wound healing in laryngeal secretions following phonotrauma: A preliminary study. Annals of Otology, Rhinology, & Laryngology. 2003; 112(12):1021–1025

[2] Branski RC, Verdolini K, Rosen CA, Hebda PA. Markers of wound healing in vocal fold secretions from patients with laryngeal pathology. Annals of Otology, Rhinology, and Laryngology. 2004; 113(1):23–29.

[3] Verdolini Abbott, K., Li, N. Y., Branski, R. C., Rosen, C. A., Grillo, E., Steinhauer, K., & Hebda, P. A. (2012). Vocal exercise may attenuate acute vocal fold inflammation. Journal of voice, 26(6), 814.e1–814.e13.

[4] Araneda OF, Salazar MP (2009) Design and evaluation of a device for collecting exhaled breath condensate. J Bras Pneumol 35:69–72

[5] Araneda OF, García C, Lagos N, Quiroga G, Cajigal J, Salazar MP, Behn C. Lung oxidative stress as related to exercise and altitude. Lipid peroxidation evidence in exhaled breath condensate: a possible predictor of acute mountain sickness. Eur J Appl Physiol. 2005 Dec;95(5-6):383-90.

[6] Araneda, O. F., Urbina-Stagno, R., Tuesta, M., Haichelis, D., Alvear, M., Salazar, M. P., & García, C. (2014). Increase of pro-oxidants with no evidence of lipid peroxidation in exhaled breath condensate after a 10-km race in non-athletes. Journal of physiology and biochemistry, 70(1), 107–115.

*The 14th International Conference on Advances in Quantitative Laryngology, Voice and Speech Research (AQL)*

# EVALUATING THE RELATIONSHIP BETWEEN RELATIVE FUNDAMENTAL FREQUENCY AND THE END OF VOCAL FOLD COLLISION IN VOICING OFFSET

**Matti D. Groll[1,2], Sean D. Peterson[3], Matías Zañartu[4], Jennifer M. Vojtech[1,2], Cara E. Stepp[1,2,5]**

[1] Department of Biomedical Engineering, Boston University, Boston, MA, USA
[2] Department of Speech, Language, and Hearing Sciences, Boston University, Boston, MA, USA
[3] Department of Mechanical and Mechanotronics Engineering, University of Waterloo, Waterloo, Ontario, CA
[4] Department of Electronic Engineering, Universidad Técnica Federico Santa María, Valparaíso, CL
[5] Department of Otolaryngology – Head and Neck Surgery, Boston University School of Medicine, Boston, MA, USA

***Keywords*: Vocal Fold Stiffness; Relative Fundamental Frequency; Voice Assessment; Laryngeal Tension**

## INTRODUCTION

Recent modeling work suggests that decreases in relative fundamental frequency (RFF) during voicing offset may be attributed to changes in vocal fold stiffness due to decreased vocal fold collision [1]. Within-speaker variability of traditional RFF measures may be caused by variable durations between the end of vocal fold collision and the final voicing cycle. This project aims to determine whether aligning RFF measures based on the last point of vocal fold collision would reduce within-speaker RFF variability.

## METHODS

A total of 45 young adult speakers with typical voices produced /ifi/ utterances with no vocal effort and maximum vocal effort during flexible high-speed videoendoscopy (HSV). RFF measures were calculated during voicing offset using two different methods: the final RFF cycle was determined by either the last point of vocal fold collision ($RFF_{COLLISION}$) or the last cycle of voicing ($RFF_{VOICING}$). HSV data were used to manually identify the end of vocal fold contact. Analyses of variance were used to determine the effects of effort and RFF method on the mean and standard deviation of RFF.

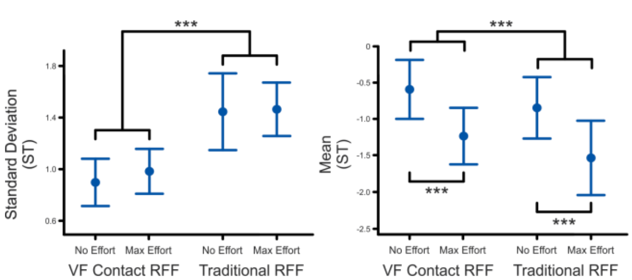



Figure 1. Average voice onset time (VOT) means and standard deviations (SDs) and 95% confidence intervals for each experimental condition. Brackets indicate significant differences between conditions ($p < .05$).

## RESULTS

Aligning by vocal fold collision ($RFF_{COLLISION}$) resulted in statistically significantly lower standard deviations ($p < .01$) than aligning by the last cycle of voicing ($RFF_{VOICING}$). RFF means were statistically higher for $RFF_{COLLISION}$ ($p = .04$); however, the degree of effort was statistically significant regardless of the method.

## DISCUSSION AND CONCLUSION

Both RFF methods demonstrate significant differences in RFF means when effort increases, as shown in previous studies, e.g. [2]. When aligning utterances based on the end of vocal fold contact during abduction, there is a decrease in within-speaker variability of RFF offset measures. This supports the theory that decreases in RFF during voicing offset are due to the reduction in system stiffness when the vocal folds cease to contact during vibration [1]. The results of this study provide important information about the physiological mechanisms behind RFF and may help improve the feasibility of RFF as an acoustic measure for estimating changes in laryngeal tension.

## ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health grants R01 DC015570, P50 DC015446, and T32 DC013017 from the National Institute on Deafness and Other Communication Disorders (NIDCD), as well as ANID grant BASAL FB0008.

## REFERENCES

[1] Serry, M. A., Stepp, C. E., & Peterson, S. D. (In Press). Physics of Phonation Offset: Towards Understanding Relative Fundamental Frequency Observations. *Journal of the Acoustical Society of America.*

[2] Lien, Y. A. S., Michener, C. M., Eadie, T. L., & Stepp, C. E. (2015). Individual monitoring of vocal effort with relative fundamental frequency: Relationships with aerodynamics and listener perception. *Journal of Speech, Language, and Hearing Research, 58*(3), 566-575.

*The 14th International Conference on Advances in Quantitative Laryngology, Voice and Speech Research (AQL)*

# MACHINE LEARNING CLASSIFIERS: STUDY TO CATEGORIZE HEALTHY AND PATHOLOGICAL VOICES

**Danilo Rangel Arruda Leite[1,2], Ronei Marcos de Moraes[1], Leonardo Wanderley Lopes[1,3]**

[1] Departamento de Estatística. Programa de Pós-graduação em Modelos de Decisão em Saúde. Universidade Federal da Paraíba, João Pessoa, Paraíba, Brasil

[2] Empresa Brasileira de Serviços Hospitalares- Ebserh, João Pessoa, Paraíba, Brasil

[3] Departamento de Fonoaudiologia. Universidade Federal da Paraíba, João Pessoa, Paraíba, Brasil

***Keywords*: Machine learning; Voice disorders; Classify pathological**

## INTRODUCTION

The voice is one of the main means of communication of the human being and its emission must be pleasant, effortless, and following the interests of the interlocutor. Any change in its emission can be classified as a voice disorder or vocal deviation [1].

Traditional methods of diagnosing voice pathologies are based on a professional's experience or expensive equipment. In this sense, computer-assisted medical systems are being increasingly used to assist professionals in diagnosing and classifying voice disorders with non-invasive methods and at a lower cost [1-2].

Upon this, this study aims to analyze different machine learning classifiers, Logistic regression (LR), Random Forest (RF), K-Nearest Neighbor (KNN), Support Vector Machine (SVM), and the combination between them (ensemble learning – EL), so that they can not only detect voice disorders but also classify healthy or pathological voices, based on cepstral measurements of the signal [3].

## METHODS

### *Participants*

The participants were defined by 305 samples of individuals (240 women and 65 men) with an average age of 36.00 ± 12.13 years were evaluated. All were seen in a higher educational institution voice laboratory between April 2012 and December 2017, and all of them signed an informed consent form. A sample set of subjects was used meeting the following criteria: adults between 18 and 65 years old; the group received a laryngological evaluation, including an otorhinolaryngological report written during the two weeks before the data collection session, to confirm the diagnosis of voice disorder; no vocal treatment (therapy or surgery) was performed before data collection.

Exclusion criteria were applied related to: patients with cognitive or neurological disorders that prevented the use of recording procedures; professional voice users and individuals who had previously received formal vocal therapy or who had undergone surgery in the head or neck region in the past 18 months; participants whose voice signals lasted less than 5 seconds, the presence of a peak cut in the acoustic signal and the signal-to-noise ratio (SNR) below 30dB SPL [4].

All computational parts were developed using the Python programming language version 3.6 [5] and the librosa library version 0.8.0 [6] to extract resources from the voice signal. Librosa is a Python library used for music and audio analysis.

### *Instrumentation and Measurement*

The vowel /Ɛ/ was selected for this study based on two reasons. First, it is an oral vowel, true medium, open and not rounded, and is considered the most average vowel in Brazilian Portuguese [7], similar to /aɛ/ in English. Second, it allows for a more neutral and intermediate position of the vocal tract and, therefore, is commonly used in visual laryngeal examination tests in Brazil.

During the extraction of the acoustic measures analyzed in this study, the samples were edited by the librosa library selecting 3 central seconds from each sustained vowel sample. This procedure aimed to exclude sections with greater variability related to the start and displacement phase of the voice.

Cepstral measurements were extracted from central 3-second samples of the sustained vowel /Ɛ/ in files of at least 6 seconds in length, with a sampling rate of 44,100 Hz. We used the MFCC technique (Mel Frequency Cepstral Coefficients) to extract the characteristics of the voice signal. The samples were edited using the librosa library, selecting 3 central seconds from each sustained vowel sample. This procedure aimed to exclude sections with greater variability related to the start and displacement phase of the voice. The phases of the proposed model are as follows: **i**. The librosa library was used to extract 3 seconds from the middle of the sustained vowel audio file /Ɛ/, using a sampling rate of 44,100 Hz. **ii**. The librosa library was used to extract resources from the MFCC, in which 20 coefficients were extracted for each signal segment, forming a vector, and returning an audio matrix to be used in the classification; **iii**. Three classification models were analyzed, as follows: Logistic regression (LR), Random Forest (RF), K-Nearest Neighbor (KNN), Support Vector Machine (SVM), and the combination of the LR, SVM, and KNN classifiers. To train the model, we divided the dataset using 3-fold cross-validation into 3 parts: 64% for the training set, 16% for validation, and 20% for testing; iv. The performance of the models was analyzed to verify the efficiency in the classification of the data, as described below.

*Analysis*

To evaluate the accuracy of the results obtained through the models, 4 measures were normally used in the scientific environment [1-2], they are: accuracy; sensitivity; specificity. These measures are related to the results of classification and true diagnosis and the ROC Curve.

The test is considered positive (deviation) or negative (healthy), and the deviation is present or absent. The test is correct when it is positive in the presence of the deviation (True Positive-VP) or negative in the absence of the deviation (True Negative-VN). In addition, the test is wrong when it is positive in the absence of the deviation (False Positive-FP), or negative when the deviation is present (False Negative-FN).

Accuracy (ACC) measures the ability of the test to correctly identify whether there is or there is not a deviation:

$$ACC = \frac{VP + VN}{VP + VN + FP + FN} \quad (1)$$

Sensitivity (SENS) is the ability of the test to correctly identify the deviation among those who have it [1-8]:

$$SENS = \frac{VP}{VP + FN} \quad (2)$$

Specificity (ESP) is the ability of the test to correctly exclude those who do not have the deviation. It is defined by the relation between the number of cases correctly classified as healthy and the total number of healthy cases [1-8]:

$$ESP = \frac{VN}{VN + FP} \quad (3)$$

## RESULTS AND DISCUSSION

The MFCC has been commonly used in the automatic classification of healthy and pathological voices and to train different types of classifiers. This technique can be considered as an approach to the structure of human auditory perception, based on human auditory behavior to extract acoustic characteristics from the voice signal [9].

The classification models LR, RF, KNN, SVM, and EL, obtained an accuracy of 94%, 96%, 95%, 77%, and 98% respectively. The ensemble learning classification model obtained the best result with an accuracy of 98%, Sensitivity of 99%, Specificity of 92%, and a ROC curve of 0.99, as can be seen in Table 1. The results suggest that the studied model performs better than individual classification models' common usage.

**Table 9. Analysis results**

| Model | Accuracy | Sensitivity | Specificity | ROC |
|---|---|---|---|---|
| EL | 98% | 98% | 92% | 99% |

## CONCLUSION

This study analyzed the machine learning models LR, RF, KNN, and EL, to classify pathological or healthy voices. The EL obtained a better result than the use of individual classification models and thus the method used can be indicated as a non-invasive tool to support the diagnosis of voice pathologies.

## REFERENCES


[1] Leite, Danilo Rangel Arruda. Moraes, Ronei Marcos. Lopes, Leonardo Wanderley. *Método de Aprendizagem de Máquina para Classificação da intensidade do desvio vocal utilizando Random Forest*. J. Health Inform, v.12, 196-201; 2020.

[2] Lopes, L. W. et al. *Evidence of Internal Consistency in the Spectrographic Analysis Protocol*. Journal of voice; 2020.

[3] Chen, L. et al. *Two-layer fuzzy multiple random forest for speech emotion recognition in human-robot interaction*. Information Sciences, 509, 150–163; 2020.

[4] Deliyski, D. D., Shaw, H. S. e Evans, M. K. *Adverse effects of environmental noise on acoustic voice quality measurements. Journal of voice* 2005; 19(1), 15–28.

[5] Python (2019). Python.org. https://www.python.org/.

[6] McFee, B. et al. *Librosa: Audio and music signal analysis in python*. In Proceedings of the 14th python in science conference 2015; p. 18–25.

[7] Gonçalves, M. I. R., Vieira, V. P. e Curcio, D. *Transfer function of Brazilian Portuguese oral vowels: a comparative acoustic analysis*. Brazilian Journal of Otorhinolaryngology, 2009; 75(5), 680–684.

[8] Moraes, R. e Martínez, L. *Computational Intelligence Applications for Data Science*; 2015; Science Direct, Out, pp. 1-2

[9] Pishgar, M., Karim, F. e Majumdar, S. *Pathological Voice Classification Using Mel-Cepstrum Vectors and Support Vector Machine*. Electrical Engineering and Systems Science; 2018.

# APPLYING ULTRASOUND IMAGING TO VISUALIZE LARYNGEAL STRUCTURES AND FUNCTIONS IN THE PEDIATRIC POPULATION

**Ip, Y.M.[1], Kwong, E.[1]**

[1] Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong

***Keywords*: Dysphonia; Pediatric Voice Disorders; Laryngeal Imaging; Ultrasound Imaging**

## INTRODUCTION

Children often show difficulties in tolerating laryngoscopic procedures. This may pose challenges to clinicians who rely on visualization of the larynx for assessment and diagnosis purpose. This study aimed to explore the application of ultrasonography as a non-invasive imaging technique to laryngeal examination in the pediatric population.

## METHODS

*Participants*

Three boys who aged 8-10 and showed no signs of puberty [1,5] and history of voice disorder participated in this study.

*Instrumentation and Measurement*

The portable ultrasound system MyLabGramma with the SL1543 appleprobe linear transducer under B-mode was used for data collection. Water gel was applied on the neck of the participants to reduce air between the two layers, in order to obtain ultrasonography with less white flash. Data exported in form of DICOM format were processed by the medical image viewer HOROS.

The participants were required to carry out both speech and non-speech tasks; namely, breath-holding, coughing, producing sustained vowels, pitch glide, glottal fry and falsetto register, and alternate production of voiceless and voiced consonants; during an ultrasonographic laryngeal examination.

*Analysis*

The data collected were analyzed qualitatively to identify the structures which are visible in ultrasonic images and characterize the manifestation of different types of laryngeal movement in the ultrasonography.

Three student speech therapists reviewed the ultrasound images and were asked to identify different tasks being performed (for both the speech and non-speech tasks); and the onset, adducted moment and offset of phonation (for the speech tasks). Data obtained in identification of different speech and non-speech tasks were analyzed by accuracy rate and the dominant error pattern of each examiner quantitatively. Data obtained in identification of events in speech tasks were analyzed by SPSS. The inter-examiner and intra-examiner reliability were analyzed using the intraclass correlation coefficient (ICC).

## RESULTS

It was found that thyroid cartilage and arytenoid cartilages, vocal folds and their mobilities could be delineated in the ultrasound images, which replicated the findings in previous studies [2,3,4,6,7]. The examiners achieved high accuracies (approximately 70% to over 90%) in identification of different speech and non-speech tasks presented in the ultrasound images. Excellent inter-examiner (ICC: 0.996) and intra-examiner (ICC: 0.996) reliabilities were obtained for the identification of phonatory events from the ultrasound images.

## DISCUSSION

Regarding the identification of different speech and non-speech tasks, qualitative analysis on the error patterns of examiners revealed that two of the examiners showed the tendency of selecting the close category distractor. It indicated that they had specific difficulty in differentiating between movements which share similar manifestations with subtle difference, for instance, identification of /s/ /z/ alternation. It is hypothesized that the manifestation of voiceless consonant /s/ in ultrasonic images are less obvious than other processes since there is only a sustained narrow gap between the vocal folds. It shares some similarities with breathing therefore could be easily ignored by the examiners and /s/ /z/ alternation could be misinterpreted as production of a sustained vowel. It could also be because of the short duration of motion in some trials.

## CONCLUSION

Ultrasound imaging may be considered as an alternative assessment tool for pediatric laryngeal examination. The excellent inter-examiner and intra-examiner reliabilities suggest further the potential of its clinical application. Future studies on dysphonic children and examiners with different background are warranted.

## REFERENCES

[1] Department of Health. (2018). *Puberty*. https://www.studenthealth.gov.hk/english/resources/resources_bl/files/puberty_parents.pdf

[2] Kristensen, M. S., Teoh, W. H., & Engelhardt, T. (2019). *Management of the Difficult Pediatric Airway*. Cambridge University Press.

[3] Ongkasuwan, J., Ocampo, E., & Tran, B. (2017). Laryngeal Ultrasound and Vocal Fold Movement in the Pediatric Cardiovascular Intensive Care Unit. *The Laryngoscope*, 127, 167-172.

[4] Rodrigues, C. T., Roriz, D., Campos, N. M. F., Macedo, J., Oliveira, C. M., Belo-Soares, P., Portilha, M. A., & Donato, P. (2019). The ultimate guide for laryngeal ultrasonography: what is there to see.

[5] Tanner, J. M. (1962). *Growth at Adolescence*. Oxford.

[6] Wang L. M., Zhu Q., Ma T., Li J.P., Hu R., Rong X.Y., Xu W., & Wang Z. C. (2011). Value of ultrasonography in diagnosis of pediatric vocal fold paralysis. *International Journal of Pediatric Otorhinolaryngology*. 75, 1186-1190.

[7] Zamudio-Burbanoa, M. A., & Casas-Arroyaveb, F.D. (2015). Airway management using ultrasound. *Colombian Journal of Anesthesiology*, *43*(4), 307-313.

*The 14th International Conference on Advances in Quantitative Laryngology, Voice and Speech Research (AQL)*

# EFFECT OF VOICE BIFURCATIONS ON ACOUSTIC AND ENDOSCOPIC MEASURES: CASE STUDIES

**Takeshi Ikuma[1,2], Lacey Adkins[1,2], Andrew McWhorter[1,2], Melda Kunduk[1,2,3]**

[1] Dept. of Otolaryngology-Head and Neck Surgery, Louisiana State University Health Sciences Center, New Orleans, USA
[2] Voice Center, The Our Lady of The Lake Regional Medical Center, Baton Rouge, USA
[3] Dept. of Communication Sciences & Disorders, Louisiana State University, Baton Rouge, USA

***Keywords*: Disordered voice; Acoustics; High-speed videoendoscopy; Voice types**

## INTRODUCTION

Disordered voice is known to change its characteristics suddenly during a phonation due to bifurcations in the underlying dynamical system (combination of laryngeal and pulmonary configurations) [1]. Using Titze's classification [2], voice signals can be categorized into three types: Type 1 (quasi-periodic), Type 2 (periodic with spurious subharmonics or modulating frequencies), and Type 3 (random). Each bifurcation can switch voice signals from one type to another or within the same type but with vastly different frequency composition. Existing objective voice measures do not explicitly consider the voice changes in disordered voice, and a typical voice assessment protocol applies a fixed-length analysis window to the middle of phonation, thus potentially including bifurcations. This presentation demonstrates the variations in objective measures when bifurcations occur for both acoustic and highspeed videoendoscopic (HSV) glottal area waveforms.

## METHODS

Four cases are sampled: (1) normal (Type I), (2) unilateral paralysis (Types I + II), (3) polyp (Types I + II), and (4) polyp (Types I + II + III). Case 2 is highlighted in this abstract. Both acoustic and HSV data were simultaneously recorded. The HSV data were then analyzed to segment the glottal pixels to form glottal area waveform (GAW). Both were then independently analyzed to identify (most prevalent if multiphonic) fundamental frequency ($f_0$) with fully automated analysis program. The selected measurements are jitter, shimmer, and cepstrum peak prominence (CPP) of Praat [3] as well as harmonic-model-based signal-to-noise ratio (SNR), H1-H2, open quotient (OQ), and relative glottal gap (RGG) [4].

## RESULTS AND DISCUSSION

Of the four cases, Case 3 best illustrates the effect of mid-phonation bifurcations. Fig.1. shows the spectrograms, $f_0$ estimates, and SNRs. The SNR is quite sensitive to voice types (for both acoustic data and GAW). The average SNRs are indicative of the pathological condition as they are substantially lower than the normal case (typically 20 to 30 dB). However, the average hides the extreme variations among voicing modes between bifurcations: Type 1 with above 20 dB SNR to the worst cases below –10 dB. This

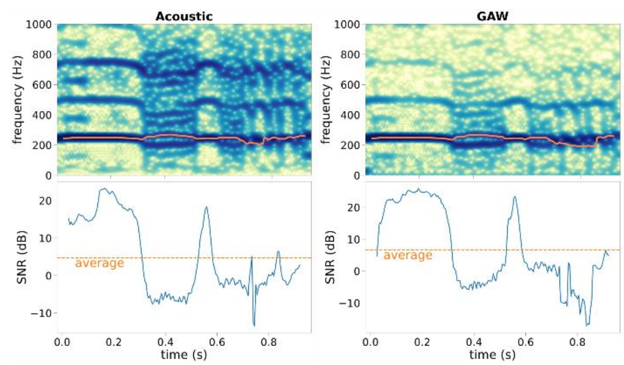



Figure 1. Spectrograms and SNRs (red line: $\boldsymbol{f_0}$ estimates).

demonstrates how inclusion of bifurcations may cause objective measures to misrepresent the effects of voice disorders. The occurrence of bifurcations is likely most detrimental in quantifying a short burst of subtle dysphonic behavior in otherwise type-1 signals (e.g., limiting Fig. 1 case to $t < 0.3$). Pre-detecting the bifurcations and focusing on dysphonic segments of phonation could improve the performance of objective parameters. In addition, the presence of mid-phonation bifurcations is related to cause intermittent voice disturbances (or pitch breaks) which could lead to perceptually worse voice quality than consistent Type-2 or Type-3 voice.

The presentation demonstrates the effects of mid-phonation bifurcations on objective measures. Results indicate the need to bifurcation-aware objective measures.

## REFERENCES

[1] I. Steinecke and H. Herzel, "Bifurcations in an asymmetric vocal-fold model," J. Acoust. Soc. Am., vol. 97, no. 3, pp. 1874–1884, 00 1995, doi: 10.1121/1.412061.

[2] I. R. Titze, Workshop on Acoustic Voice Analysis: Summary Statement. Denver, CO: National Center for Voice and Speech, 1994. [Online]. Available: http://www.ncvs.org/freebooks/summary-statement.pdf

[3] P. Boersma and D. Weenink, Praat: doing phonetics by computer. 2021.

[4] T. Ikuma, M. Kunduk, and A. J. McWhorter, "Objective quantification of pre and post phonosurgery vocal fold vibratory characteristics using high-speed videoendoscopy and a harmonic waveform model," J. Speech Lang. Hear. Res., vol. 57, pp. 743–757, Jun. 2014, doi: 10.1044/2013_JSLHR-S-12-0202.

# SELF-INFORMED INSTRUMENTS APPLIED IN PATIENTS WITH UNILATERAL VOCAL CORD PARALYSIS: A SYSTEMATIC REVIEW OF THE 2010-2020 DECADE

**Jenny F. Cárdenas M** [1]

[1] Universidad Nacional de Colombia, Bogotá, Colombia

***Keywords*: Voice; Voice outcome survey; Unilateral vocal cord paralysis; Quality of life; self-reported**

## INTRODUCTION

The Voice Outcome Survey (VOS) is the first validated self-report instrument that was developed to measure and evaluate the outcomes of treatments related to quality of life in people with *unilateral vocal cord paralysis* (UVCP) [1]. It evaluates aspects such as phonation, swallowing, social functioning and breathing [2]. It was used in various types of treatments, translated, and validated in different countries. However, the search for studies in the last decade was limited in terms of information on adaptation to other types of cultures, languages or used to assess outcomes after any treatment. Unfortunately, VOS does not provide reliable data for individual decision-making as it is a short instrument and due to its lack of versatility, but it could serve as the basis for creating a more comprehensive, psychometrically reliable, patient-based disease-specific instrument, validated and clinically applicable [3]. Therefore, information is required to provide evidence on the application of a self-report instrument for the specific population of patients with UVCP. As a consequence of the disuse of VOS from his latest research in 2007, the question arises: What is the most used quality of life self-report instruments in patients with UVCP 2010-2020? And in the instruments found, ¿Which of these meets the criteria of reliability, validity and detect changes reported after different post-test and pre-test treatments?

## METHODS

A systematic review of the last decade 2010-2020 was carried out to identify the most used self-reporting instruments in patients with UVCP in published studies. The review considered some inclusion criteria in the extracted articles, such as the use of the pre-test and post-test self-reporting instruments in everything related to (UVCP), measuring the effect of the change due to the treatments. The databases used, exclusion criteria and the article selection process can be seen in Figure 1.

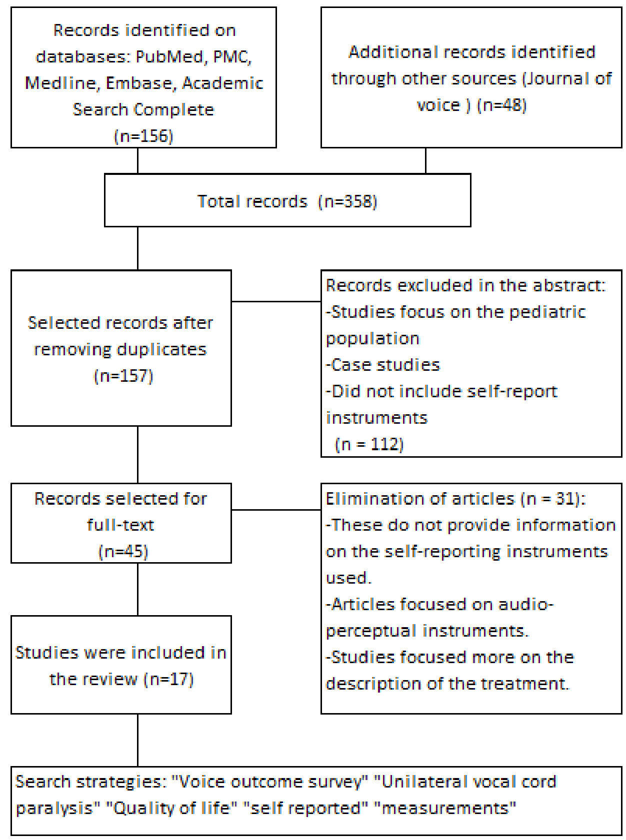



Figure 1. Summary of the systematic review methodology

## RESULTS AND DISCUSSION

In total, 17 studies with different intervention approaches and different types of treatment were selected for UVCP: From the latest publications of studies on self-report instruments applied to our specific pathology, it was found that in 2019 a study on self-assessment tools of Voice results in (UVCP) patients undergoing arytenoid adduction (AA) used the vocal disability survey (VHI-30) [4]. In 2018, the VOIs (Voice Outcome Indicators) used for the evaluation of surgical treatments in UVCP were analyzed, surgeons indicate that they prefer (VHI-30) [5]. The same year, research was conducted to identify which VOIs are used most frequently, and they are more relevant in terms of significant changes; the author states that the VHI-30 was chosen because it is well known, commonly used, and provides a greater range of potential scores than the shorter versions [6]. VHI-30 has the largest number of publications on validity and response to change [7]. Nevertheless, for the year 2017 a study was found where they designed their own instrument, Thyroidectomy-related voice questionnaire (TVQ) is a self-assessment instrument developed to measure voice quality after thyroid surgery in patients with UVCP [8]. Subsequently, in 2016 a critical review of the literature was found, which compares the interventional approaches for (UVCP), the use of self-reported instruments that included (VHI-30) is evidenced to obtain postoperative scores in medialization thyroplasty compared to injection laryngoplasty [9].

Furthermore, it was found that VHI-30 was adapted to a German version of 12 questions for self-assessment [10]. Also, it was used for the treatment of Injection of the vocal cords with hyaluronic acid guided by laryngeal electromyography for UVCP [11]. Additionally, it is evidenced that it was used in the measurements of significant pre and postoperative changes to evaluate UVCP surgical treatments [12]. Self-perception tools were used to assess the impact on the quality of life of the patients, the use of the VHI-30 tool being more common. The results indicate that VHI-30 had a high "percentage of significance" compared to other tools [13]. Therefore, due to the greater number of publications, many authors prefer to use it because they consider it “more reliable”, but the content validity of the VHI-30 may be overestimated, and it is important to evaluate the swallowing and breathing aspects that are directly affected in patients with UVCP [14].

## CONCLUSION

The quality-of-life self-report instrument to evaluate voice results most used in patients with UVCP in the decade 2010-2020 is the VHI-30 because it meets criteria of reliability, validity and compares the preoperative and postoperative voices in patients undergoing various treatments. VHI replaced the VOS tool that was originally created for this condition. However, there are those who do not settle for none and create new tools to evaluate each specific treatment; Therefore, it is necessary to continue in the search and creation of a specific tool for the disease and not for the treatment, psychometrically reliable, validated, clinically applicable and generalizable for vocal disability attributable to UVCP; This makes it possible to assess the disability of affected patients in all laryngeal, cultural and psychosocial domains in order to be more empathetic, communicate more effectively, and better personalize treatment plans, generating better care.

## ACKNOWLEDGMENTS

To the professor / speech language pathology at the Universidad Nacional de Colombia, Luisa Fernanda Angel Gordillo, for her contribution during the research of this article, providing technical assistance, critical review of the manuscript, suggestions, and ideas.

## REFERENCES

[1] Gliklich RE, G. R. (1999) Validation of a voice outcome survey for unilateral vocal cord paralysis. *Otolaryngol Head Neck Surg*; 153-158

[2] Fang, T. L. (2007). Assessment of Chineseversion voice outcome survey in patients. *Otolaryngol Head Neck Surg*, 136 (5), 752-756.

[3] Franic, D. B. (2005). Psychometric Evaluation of Disease Specific Quality of Life. *Journal of Voice*, 300-315.

[4] Watanabe K, S. T.-H. (2019). Characteristics of the Voice Handicap Index for Patients with Unilateral Vocal Fold Paralysis Who Underwent Arytenoid Adduction. *Journel of voice*, 2019; S08921997(18)30482-X

[5] Desuter, G, M. D.-B. (2019). Voice outcome indicators for unilateral vocal fold paralysis surgery: A survey among surgeons. European Annals of Otorhinolaryngology, *Head and Neck Diseases*, 343-347.

[6] Francis, D. S. (2018). Life Experience of Patients with Unilateral Vocal Fold Paralysis. JAMA otolaryngology-- head & neck surgery, 144(5), 433–439.

[7] Walton, C. C. (2018). Voice Outcome Measures for Adult Patients with Unilateral Vocal. The Laryngoscope. doi: 10.1002/lary.27434

[8] Choi YS, J. Y. (2017). Factors Predicting the Recovery of Unilateral Vocal Fold Paralysis After Thyroidectomy. World J Surg., 42(7):2117-2122.

[9] Siu J, T. S. (2016). A comparison of outcomes in interventions for unilateral vocal fold paralysis: A systematic review. Laryngoscope; 126(7):1616-1624.

[10] Koelmel, J., & Sittel, C. (2014). Stimm- und Lebensqualität nach Injektionslaryngoplastik mit VOXImplants (Polidimetilsiloxano). LaryngoRhino-Otologie, 93 (05): 316 – 320

[11] Wang CC, C. M. (2012). Laryngeal electromyography-guided hyaluronic acid vocal fold injection for unilateral vocal fold paralysis—preliminary results. J. Voice;26(4):506-514

[12] Van Ardenne N, V. J. (2011). Medialization thyroplasty: vocal outcome of silicone and titanium implant. Eur Arch Otorhinolaryngol, 268(1):101-107.

[13] Desuter G, D. M.-B. (2018). Voice outcome indicators for unilateral vocal fold paralysis surgery: a review of the literature. Eur Arch Otorhinolaryngol., 275(2):459-468.

[14] Francis, D. D. (2016). Voice-Related Patient-Reported Outcome Measures: A Systematic. Journal of Speech Language and Hearing Research, 60 (1).

# WAVELET PACKET TRANSFORM AND MULTILAYER PERCEPTRON TO IDENTIFY VOICES WITH A MILD DEGREE DEVIATION

**Mateus Morikawa[1], Danilo Hernane Spatti[2], Maria Eugenia Dajer[1]**

[1] Department of Electrical Engineering, Federal University of Technology – Paraná (UTFPR-CP), Brazil.
[2] Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, SP, Brazil.

***Keywords*: Voice; Artificial Neural Networks; Wavelet Packet Transform; Multilayer Perceptron.**

## INTRODUCTION

The voice is one of the main tools of human communication. According to [1] voice is basically produced by three processes: movement of the vocal folds interrupting the subglottic airflow, followed by the resonance and articulation of this fundamental sound, which takes place in the supraglottic vocal tract. Any change in this complex mechanism can mean a change in the vocal quality of a subject. As human voice is essentially an auditory-perceptual signal, any voice disorder is usually recognized as a deviation in vocal quality [2]

Since biological signals, as voice, are not stationary, the application of the Fourier transform does not prove to be an accurate alternative to perform an acoustical analysis. However, the Wavelet Packet Transform (TWP) has been used as an alternative tool, acting as an extractor of signal characteristics [3]. In addition, another tool, such as Artificial Neural Networks (ANNs) can improve the performance of pattern classification in voice signals.

The purpose of this paper is to study an alternative way to identify voices with a mild degree deviation, using the Wavelet Packet Transform and Artificial Neural Networks.

## METHODS

*Database*

The database consisted of 74 audio files classified into 3 groups: 25 voices without vocal deviation, 29 voices with mild vocal deviation, and 20 voices with moderate vocal deviation, according to the perceptual-auditory indices observed [4]. The database was provided by Dr. Fabiana Zambóm and further details of the data collection and the classification can be found at [5].

Since the goal of this paper is to identify voices with a mild degree of deviation, we divided the data set into two groups: G1 = voices with a mild degree of deviation and G2 = voices without deviation and voices with a moderate degree of deviation.

*Procedures*

For this work, we use MATLAB (student license), and the procedures were composed of the following steps: pre-processing, segmentation, characteristic extraction, classification, and post-processing.

The pre-processing step consisted of removing periods of silence from audio files, as well as any types of sound that are not of the patient, called artifacts.

In the segmentation step, the objective was to separate the data into a set of training (80%) and a set of testing (20%). For each voice signal, a window of 4096 discretized samples and 50% overlap was applied. Table 1 shows the number of samples for training and testing in groups 1 and 2, before and after segmentation.

**Table 10. Number of samples for Groups (G1 and G2), pre and post segmentation.**

| | Pre-segmentation | | Post-segmentation | |
|---|---|---|---|---|
| **Register** | **G1** | **G2** | **G1** | **G2** |
| Training | 23 | 36 | 4402 | 7723 |
| Test | 6 | 9 | 1156 | 1843 |

For the extraction of characteristics step, the Wavelet Packet (TWP) transform was used, since this one obtains information in both, the domain of time and frequency. We used the Daubechies 2 and Symlet 2 families for extracting the energy and Shannon's entropy values from the coefficients of approximation and detail.

The processing step was performed with the Multilayer Perceptron (MLP) network with the Levenberg-Marquardt learning algorithm [6], using the hyperbolic tangent function in the intermediate layers, and a learning rate of 0.2. The topology used is represented by two intermediate layers, which have 1 neuron in the first and 2 neurons in the second layer. Since the MLP uses a supervised learning process, it is necessary to indicate the desired values of the answers. Thus, the output has defined the vector [1 -1] for the class Group 1. To the samples of Group 2, the vector [-1 1] was defined. If the result did not fit into either option, the designated vector was [2 2] indicating uncertainty.

Finally, the post-processing step consisted of adjusting the output vectors produced by MLP. Therefore, it has been established a 98% degree of reliability. Thus, each of the two positions of the output vector was compared to the threshold of ± 0.98. Therefore, if the term value was higher than 0.98, this would receive value 1. If the term value was less than -0.98, this would receive -1. For values between -0.98 and 0.98, the term would receive 2.

## RESULTS

To prevent the randomization of the initialization of synaptic weights from interfering in the final answer, the network was trained and tested 10 times. Aiming to carry out a more detailed analysis of the classifier, the confusion matrices of each wavelet family were assembled with the average of the 10 tests.

According to Tables 2, 3, 4, and 5, it is possible to observe that the proposed classification algorithm obtained an accuracy of 99.76% and 99.56% for energy and entropy measures using the Symlet 2 family, and 91.17% and 70.01% for the same measures using the Daubechies 2 family.

**Table 2. Confusion matrix with accuracy percentage using Symlet 2 family and energy values.**

| | G1 | G2 | Uncertainty |
|---|---|---|---|
| G1 | 99,75 % | 0,15% | 0,10% |
| G2 | 1,14% | 97,57% | 1,29% |

**Table 3. Confusion matrix with accuracy percentage using Symlet 2 family and entropy values.**

| | G1 | G2 | Uncertainty |
|---|---|---|---|
| G1 | 99,56 % | 0,31% | 0,13% |
| G2 | 2,19% | 96,29% | 1,52% |

**Table 4. Confusion matrix with accuracy percentage using Daubechies 2 family and energy values.**

| | G1 | G2 | Uncertainty |
|---|---|---|---|
| G1 | 91,17 % | 3,68% | 5,15% |
| G2 | 0,50% | 98,29% | 1,21% |

**Table 5. Confusion matrix with accuracy percentage using Daubechies 2 family and entropy values.**

| | G1 | G2 | Uncertainty |
|---|---|---|---|
| G1 | 70,01 % | 1,97% | 28,02% |
| G2 | 0,34% | 86,75% | 12,91% |

The results presented in the confusion matrices suggest that Symlet 2 outperformed Daubechies 2, as can be seen from the measures of uncertainties, errors, and successes in identifying the desired class. Although previous work has shown that the Daubechies 2 families and Symlet 2 were efficient for the analysis of vocal signals, for this study the performance of Daubechies 2 using entropy did not have a good result.

## CONCLUSION

This research aimed to train a neural network specialist in recognizing mild voice disorders. It is concluded that the MLP proved to be robust enough to generate a high rate of correctness in its classification, which, in most cases, surpassed 99% accuracy with 98% reliability.

Also, it is observed that only 3 neurons in the intermediate layers have already been enough to perform a good generalization, not requiring thus, a great computational performance.

Future work will explore the use of other wavelet families and the use of larger databases.

## ACKNOWLEDGMENTS

Special thanks to Dr. Fabiana Zambón for sharing the voice database.

## REFERENCES

[1] Imamura, R. Tsuji, DH; Sennes, LU. Fisiologia da laringe. IN: Pinho, SMR; Tsuji, DH, Bahadana, SC. Fundamentos de Laringologia e Voz. Rio de Janeiro. Revinter Ltda. 2006.

[2] Behlau, M.; Rocha, B.; Englert, M.; Madazio, G. Validation of the Brazilian Portuguese CAPE-V Instrument—Br CAPE-V for Auditory-Perceptual Analysis, Journal of Voice, 2020.

[3] Lima, AAM.; De Barros, FKH.; Yoshizumi, VH; Spatti, DH; Dajer, ME. Optimized Artificial Neural Network for Biosignals Classification Using Genetic Algorithm. *Journal Of Control, Automation and Electrical Systems*, V.30, pages371–379, 2019.

[4] Yamasaki, R. et al. Auditory-perceptual Evaluation of Normal and Dysphonic Voices Using the Voice Deviation Scale. Journal of Voice, v. 31, n. 1, p. 67-71, 2016.

[5] Zambon, F.C. Estratégias de enfrentamento em professores com queixa de voz. São Paulo, 2011.

[6] Silva, IND; Spatti, DH.; Flauzino, RA. Redes Neurais Artificiais para engenharia e ciências aplicadas. São Paulo, SP: Artliber, 2010.

# INFLUENCE OF A POSTERIOR GLOTTAL OPENING ON DISSIPATED COLLISION POWER: SYNTHETIC SELF-OSCILLATING VOCAL FOLD INVESTIGATIONS

**Mohsen Motie-Shirazi[1], Sean D. Peterson[2], Matías Zañartu[3], Byron D. Erath[1]**

[1] Department of Mechanical and Aeronautical Engineering, Clarkson University, Potsdam, NY, USA
[2] Department of Mechanical and Mechatronics Engineering, University of Waterloo, Waterloo, ON, CA
[3] Department of Electronic Engineering, Universidad Técnica Federico Santa María, Valparaíso, Chile

***Keywords*: Posterior Glottal Opening, Vocal Fold Collision, Dissipated Power, Vocal Fold Damage**

## INTRODUCTION

A posterior glottal opening (PGO) is a common physiological deficiency that occurs because of incomplete closure of the vocal folds (VFs) at the posterior glottal chink. The presence of a PGO is associated with the production of a breathy voice with a lower sound pressure level (SPL) [1]. In response, an individual commonly implements physiological behaviors to increase the output SPL, such as by raising the subglottal pressure. This puts the VFs at risk of increased contact pressure and subsequently increased dissipated collision power during contact, which are believed to contribute to the development of benign VF lesions [2, 3].

Surprisingly, there is limited knowledge about the effect of a PGO and the associated compensatory actions on VF contact mechanics. Direct measurements of in vivo contact pressure are highly challenging due to the invasive nature of these experiments. Therefore, investigations into the effects of a PGO have been limited to lumped-element [4, 5] and computational [6] modeling. Unfortunately, a shortcoming of these models is that the collision forces must be prescribed a priori, which undermines the accuracy of these methods.

The objective of this study is to quantify the influence of compensatory behaviors on VF contact pressure and dissipated collision power as a function of PGO area, thereby investigating the likelihood of VF damage. For this purpose, self-oscillating synthetic VF models are employed. Understanding this effect will yield insight into the pathophysiology of VF phonotrauma and will be helpful in the diagnosis and treatment of these voice disorders.

## METHODS

The experiments were performed by using a four-layer synthetic silicone VF model in a hemilaryngeal flow facility. The details of the VF model geometry and mechanical properties can be found in prior work [7].

In this flow facility, pressurized air enters a 0.3 $m^3$ plenum chamber, representing the lung volume, and exits the chamber through a 15.0 cm long tracheal channel with a rectangular cross-sectional area of 213.0 $mm^2$. A Dwyer RMC 103-SSV flow meter measures the flow rate, and a Kulite ET-3DC measures the subglottal pressure at a distance of 30.0 mm before the tracheal channel exit. A silicone VF model is mounted at the end of the subglottal tract, which oscillates against a solid hemilaryngeal plate. A simplified model of the vocal tract mimicking the human vocal tract geometry when producing the vowel /o/ is placed at the VF exit. A channel with a semi-circular cross-sectional area is included posterior to the VF model, which connects the subglottal and supraglottal tracts and creates a bypass, representing a PGO. Four different PGO areas of 2, 5, 8, and 10 $mm^2$ were investigated in this study.

The contact pressure was measured with a Millar Mikro-Cath pressure sensor embedded in the hemilaryngeal plate at the mid anterior-posterior direction. The hemilaryngeal plate is connected to a linear slide to adjust the location of the pressure sensor with an accuracy of 0.0254 mm in the inferior-superior direction. The SPL is measured with a B&K 4189 microphone, and the VF oscillations are recorded with a Photron AX200 high-speed camera at 80,000 frames per second.

For each PGO area, the subglottal pressure was adjusted such that the VF produced a target SPL of 88.0 dB. The contact pressure distribution within the inferior-superior length of contact was measured, and the maximum value of contact pressure ($p_{\max}$) was found. The dissipated collision power, $W_d$, is found by computing the difference between the kinetic power of the VF prior to contact, $W_k$, and the power restored to the VF during the contact phase, $W_c$. The kinetic and contact power were computed by.

$$W_k = \frac{0.06}{T} \rho v_c^2 \tag{1}$$

$$W_c = \frac{p_{avg} l_c \delta_c}{AT} \tag{2}$$

where $\rho$ is the VF density, $v_c$ is the medial surface velocity of the VF immediately preceding contact, $T$ is the oscillation period, $p_{avg}$ is the average contact pressure along the inferior-superior direction, $l_c$ is the inferior-superior length of contact, $A$ is the coronal cross-sectional area of the VF, and $\delta_c$ is the fictitious penetration depth of the VF during contact, estimated by assuming a Hertzian model of contact. The derivations of the governing equations above are not presented here for brevity. Based on (1) and (2), the dissipated power includes the effects of not only the contact pressure distribution but also the surface velocity of the VF. Therefore, the dissipated power can be a more comprehensive representation of VF damage than measures of only the maximum contact pressure.

## RESULTS

It was found SPL decreases as PGO area increases. Therefore, a higher subglottal pressure was required to compensate for the reduced SPL. Raising the subglottal pressure resulted in a nonlinear increase in the maximum contact pressure. The compensated subglottal pressure, $p_{sub}$, and the associated maximum contact pressure, $p_{max}$, are plotted as a function of PGO area, $A_{PGO}$, in Fig.1(a). A modest increase of 5% in the subglottal pressure when the PGO area increased to 10 $mm^2$ resulted in a more remarkable increase of 12% in the maximum contact pressure. A similar trend was observed in the dissipated power magnitude, $W_d$, as a function of PGO area, shown in Fig.1(b). Interestingly, although the maximum contact pressure increased by only 12% for the largest PGO area, the dissipated power increased by 122%. This was because the surface velocity of the VF increased as well, which resulted in higher kinetic power and subsequently, higher dissipated power.

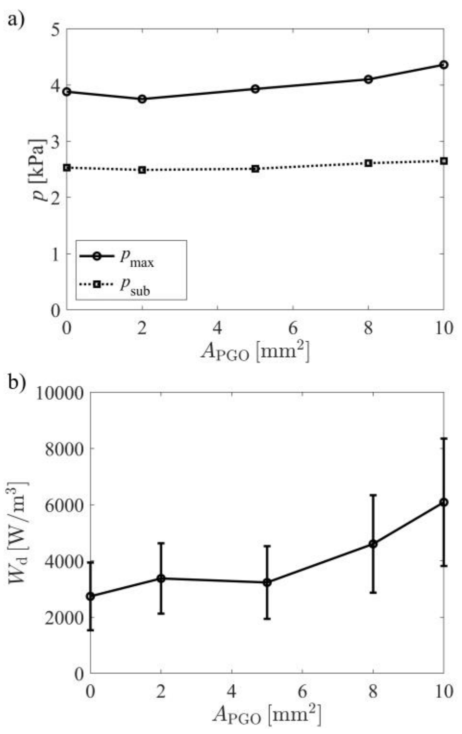



Figure 1. The (a) subglottal pressure ($p_{sub}$) and peak contact pressure ($p_{max}$), and (b) dissipated collision power as a function of posterior glottal opening area, where the subglottal pressure was adjusted to achieve a target SPL of 88.0 dB.

## DISCUSSION

The marked increase in dissipated power, even though the peak contact pressure change was modest, predicts that the likelihood of phonotrauma greatly increases with increasing PGO area. Furthermore, only measuring maximum contact pressure may not be sufficient to adequately assessing the risk of phonotrauma. The advantage of using dissipated power to evaluate VF damage is that it considers both the spatial distribution of contact pressure as well as the VF contact velocity. This provides new insights into the phonotraumatic consequences of compensatory behaviors.

## CONCLUSION

Compensating for reduced SPL due to the presence of a PGO resulted in a nonlinear increase in the maximum contact pressure and dissipated power. The dissipated power had a more significant increase, indicating that it may provide a more accurate metric in assessing VF damage.

## ACKNOWLEDGMENTS

This research was supported by the National Institutes of Health (NIH) National Institute on Deafness and other Communication Disorders Grant P50 DC015446.

## REFERENCES

[1] Schneider B, Bigenzahn W. "Influence of glottal closure configuration on vocal efficacy in young normal-speaking women." *J Voice* 17.4 (2003): 468-480.

[2] Titze, IR, and Hunter EJ. "Comparison of vocal vibration-dose measures for potential-damage risk criteria." *J Speech Lang Hear Res* 58.5 (2015): 1425-1439.

[3] Hillman RE, Stepp CE, Van Stan JH, Zañartu M, Mehta DD. "An updated theoretical framework for vocal hyperfunction." *Am J Speech Lang Pathol* 29.4 (2020): 2254-2260.

[4] Zañartu M, Galindo GE, Erath BD, Peterson SD, Wodicka GR, Hillman RE. "Modeling the effects of a posterior glottal opening on vocal fold dynamics with implications for vocal hyperfunction." *J Acoust Soc Am* 136.6 (2014): 3262-3271.

[5] Galindo GE, Peterson SD, Erath BD, Castro C, Hillman RE, Zañartu M. "Modeling the pathophysiology of phonotraumatic vocal hyperfunction with a triangular glottal model of the vocal folds." *J Speech Lang Hear Res* 60.9 (2017): 2452-2471.

[6] Zhang Z. "Cause-effect relationship between vocal fold physiology and voice production in a three-dimensional phonation model." *J Acoust Soc Am* 139.4 (2016): 1493-1507.

[7] Motie-Shirazi M, Zañartu M, Peterson SD, Mehta DD, Kobler JB, Hillman RE, Erath BD. "Toward development of a vocal fold contact pressure probe: sensor characterization and validation using synthetic vocal fold models." *Appl Sci* 9.15 (2019): 3002.

# INVESTIGATING ANTAGONISTIC MUSCLE CONTROL IN NON-PHONOTRAUMATIC VOCAL HYPERFUNCTION USING A TRIANGULAR BODY-COVER MODEL OF THE VOCAL FOLDS

**Gabriel A. Alzamendi[1], Sean D. Peterson[2], Byron D. Erath[3], Robert E. Hillman[4], Matías Zañartu[5]**

[1] Institute for Research and Development on Bioengineering and Bioinformatics CONICET-UNER, Oro Verde, Argentina
[2] Mechanical and Mechatronics Engineering, University of Waterloo, Waterloo, Ontario, Canada
[3] Department of Mechanical and Aeronautical Engineering, Clarkson University, Potsdam, New York, USA
[4] Center for Laryngeal Surgery and Voice Rehabilitation, Massachusetts General Hospital, Boston, Massachusetts, USA
[5] Department of Electronic Engineering, Universidad Técnica Federico Santa María, Valparaíso, Chile

***Keywords*: model of glottal function, non-phonotraumatic hyperfunctional phonation, triangular body-cover model, intrinsic laryngeal muscles**

## INTRODUCTION

Non-phonotraumatic vocal hyperfunction (NPVH) is a common condition that causes dysphonia without vocal fold (VF) tissue trauma. It is attributed to aberrant control of laryngeal muscles which can create high levels of stiffness and tension in the VFs. Despite its significant prevalence, very little is known about the underlying physical mechanisms that contribute to NPVH. In this study we implement a numerical model of voice production to investigate the role that aberrant control of antagonistic laryngeal muscles may play in NPVH.

## METHODS

*Physiological muscle-controlled model of phonation*

A voice production model was implemented for simulating human phonation through the independent activation of all five intrinsic laryngeal muscles. The model includes a biomechanical scheme for the muscle control of phonatory posturing [1], a triangular body-cover VF structure [2], a set of biomechanical rules for controlling the oscillatory VF model [3], an anatomically- relevant glottis shape, and a smooth zipper-like contact. Voice simulation also includes the three-way fluid-structure-acoustics interaction at the glottis, as well as acoustic wave propagation through the subglottal and supraglottal tracts.

*Simulation of NPVH*

NPVH was simulated as higher-than-normal imbalanced activations of intrinsic muscles while maintaining a given prephonatory laryngeal posture. Sustained vowel and dynamic /vcv/ were simulated. The activation scenarios addressed the antagonist role of intrinsic muscles on both the phonatory laryngeal posturing and VF biomechanical configuration. Performance of phonation was described by assessing the differences on the glottal aerodynamics, vocal fold oscillations, and voice spectral content.

## RESULTS

The simulations with sustained vowels showed that posturing scenarios with increased muscle activation yield elevated open quotients and subglottal pressures when fundamental frequency and sound pressure level are fixed, which is in agreement with previous clinical observations. Simulations with dynamic conditions better illustrated changes in vocal onset, transient components, and voice spectral content between the NPVH and control conditions.

## DISCUSSION AND CONCLUSIONS

Applying a biomechanical muscle control of the larynx in a physiological voice simulator aids to provide new insights into the pathophysiological mechanisms of atypical vocal function. The proposed framework allows to investigate the acoustic, aerodynamic, and vibratory effects of heightened and unbalanced muscle activation associated with NPVH. Future efforts will be devoted to simulating relative fundamental frequency with this framework.

## 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] I. R. Titze and E. J. Hunter, "A two-dimensional biomechanical model of vocal fold posturing," J. Acoust. Soc. Am., vol. 121, no. 4, pp. 2254–2260, 2007.

[2] G. E. Galindo, S. D. Peterson, B. D. Erath, C. Castro, R. E. Hillman, and M. Zañartu, "Modeling the Pathophysiology of Phonotraumatic Vocal Hyperfunction With a Triangular Glottal Model of the Vocal Folds," J. Speech, Lang. Hear. Res., vol. 60, no. 9, pp. 2452–2471, 2017.

[3] I. R. Titze and B. H. Story, "Rules for controlling low-dimensional vocal fold models with muscle activation," J. Acoust. Soc. Am., vol. 112, no. 3, pp. 1064–1076, 2002.

# A COMPUTATIONAL STUDY ON THE IMPLANT SHAPE OPTIMIZATION FOR TREATMENT OF UNILATERAL VOCAL FOLD PARALYSIS

**Mohammadreza Movahhedi[1], Biao Geng[1], Qian Xue[1], Xudong Zheng[1*]**

[1] Department of Mechanical Engineering, University of Maine, Orono, Maine, USA

***Keywords*: Numerical Simulation; Vocal Fold Paralysis; Type I Medialization; Implant Optimization**

## INTRODUCTION

The most common permanent procedure for treating unilateral vocal fold paralysis/paresis (UVFP) is medialization laryngoplasty [1], which restores the vocal fold vibrations by implanting a configured support structure to the paretic fold to reduce the glottal gap during phonation. The optimal voice outcomes depend upon the exact placement of the implant relative to the position of the underlying vocal fold, specifically the shape and position of the implant. Suboptimal voice outcomes and high revision rates reflect the significant challenges inherent in this procedure. In this study, a computational framework is developed for searching for the shape of the implant that produces the optimal aerodynamic and acoustic outcomes of the medialization procedure. The algorithm combines a genetic algorithm (GA) based optimization program with a patient-specific larynx computer model, which simulates the entire phonation process from vocal fold posturing to flow-structure-acoustic interaction (FSAI) as well as virtual surgery of implant insertion.

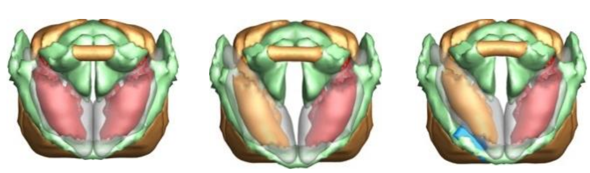



Figure 3. Vocal fold pre-phonatory configurations under normal condition, UVFP and surgical correction.

## METHODS

A three-dimensional high-fidelity larynx model was previously developed to simulate vocal fold posturing with different combination of intrinsic laryngeal muscle activations [2]. This model is further utilized to generate the pre-phonatory configurations of the vocal folds under normal, UVFP and surgical correction conditions. The UVFP condition is modeled by deactivating all the intrinsic muscles from the right side. The medialization procedure is modeled by virtually inserting a trapezoidal prism-shaped implant into the paralyzed fold. Figure 1 shows the simulated three corresponding pre-phonatory configurations. The GA based optimization solver is further coupled with the FSAI simulations to optimize the implant shape, characterized by three parameters: the insertion depth, anterior-posterior angle, and inferior-superior angle, for the desired aerodynamic and acoustic outcomes. In this preliminary study, two different objective functions are used for the optimization: (1) the aerodynamic function including the maximum flow deceleration rate (MFDR) and flow leakage and (2) the acoustic function including the cepstral peak prominence (CPP) and sound intensity.

## RESULTS AND DISCUSSION

For both objective functions, the GA algorithm is successfully converged after the 7th generation with a 63 population. To show the effect of the optimized implant on improving phonation, the aerodynamic and acoustic features of the healthy, UVFP, and two optimized medialization cases are compared in Table 1. The $Imp_{Aero}$ and $Imp_{Ac}$ represent the implants based on the aerodynamic and acoustic objective functions, respectively. The simulations show that both the aerodynamic and acoustic features of healthy case are well within the typical range of normal phonation [3-5]. The UVFP yields low MEDR, high leakage, low CPP and low sound intensity. Due to the high level of noise, HNR is not measurable under UVFP. Both medialization cases can restore the aerodynamic and acoustic quantities to a level close to the healthy case. For example, the MFDR, CPP and sound intensity in the medialization cases are close or even better than those in the health case, although the leakage flow in the medialization cases still remains relatively higher than that in the healthy case. The insertion of the implant does not affect the frequency. Comparing between $Imp_{Aero}$ and $Imp_{Ac}$, $Imp_{Ac}$ yields a better acoustic outcome with higher CPP, HNR and sound intensity.

**Table 11. Aerodynamics and acoustic features of all cases.**

| Feature | Healthy | UVFP | $Imp_{Aero}$ | $Imp_{Ac}$ |
|---|---|---|---|---|
| F0 [Hz] | 177 | 178 | 168 | 165 |
| MFDR [ml/ms$^2$] | 160 | 44 | 195 | 187 |
| Leakage [ml/s] | 12 | 570 | 23 | 45 |
| CPP | 24 | 14 | 21 | 29 |
| HNR [dB] | 23.5 | N/A | 15.1 | 28 |
| Intensity [dB] | 81.5 | 62.0 | 77.5 | 81.4 |

### CONCLUSION

A computational framework was successfully developed to optimize the implant shape for Thyroplasty Type 1 phonosurgery. The implant was successfully optimized, and the voice outcomes are significantly improved using both aerodynamic and acoustic criteria.

### ACKNOWLEDGMENTS

This work is supported by the National Science Foundation under Grant #1652632.

### REFERENCES

[1] Isshiki N, Okamura H, Ishikawa, T. Thyroplasty type I (lateral compression) for dysphonia due to vocal cord paralysis or atrophy. *Acta oto-laryngologica*. 1975;80(1-6); 465-473.

[2] Geng B, Pham N, Xue Q, Zheng X. A three-dimensional vocal fold posturing model based on muscle mechanics and magnetic resonance imaging of a canine larynx. *The Journal of the Acoustical Society of America*. 2020;147(4):2597-608.

[3] Peterson, G.E. and Barney, H.L., 1952. Control methods used in a study of the vowels. *The Journal of the acoustical society of America*, 1952;24(2):175-184.

[4] Teixeira JP, Fernandes PO. Jitter, Shimmer and HNR classification within gender, tones and vowels in healthy voices. *Procedia technology*. 2014; 16:1228-37.

[5] Brockmann-Bauser, M., Van Stan, J.H., Sampaio, M.C., Bohlender, J.E., Hillman, R.E. and Mehta, D.D. Effects of vocal intensity and fundamental frequency on cepstral peak prominence in patients with voice disorders and vocally healthy controls. *Journal of Voice.* 2019.

# TESTING THE VALIDITY OF THE QUASI-STEADY ASSUMPTION IN VOCAL FOLD VIBRATION

**Xiaojian Wang[1], Anil Palaparthi[2], Ingo R. Titze[2], Xudong Zheng[1], Qian Xue[1]**

[1] Department of Mechanical Engineering, University of Maine, Orono, Maine, USA

[2] The National Center for Voice and Speech, University of Utah, Salt Lake City, Utah, USA

***Keywords*: Quasi-steady Assumption; Numerical Modeling; Fluid-structure Interaction**

## INTRODUCTION

The quasi-steady flow assumption is that the inertial properties of the air in the glottis and the effect of rapidly moving boundaries have negligible effects on the pressure in the glottis, even if vorticity and jet formation are present. The validity of this assumption is of vital importance for both physical and computational modeling of voice production. Although it has been verified for simplified unchanging glottal shapes [1-4], the quasi-steady assumption has not yet been validated for complex and cyclically changing glottal geometries. In addition, previous studies mostly focused on the normal speaking fundamental frequency (of the order of 100 Hz), the legitimacy of the quasi-steady assumption at higher fundamental frequencies has not been well understood. This work was aimed to use numerical method to further investigate the range of validity of the quasi-steady assumption for physiologically more realistic glottal shapes. The error in glottal flow and wall pressure caused by applying the assumption were quantified and assessed at different fundamental frequencies.

## METHODS

Two normal modes of vocal fold vibration, each of which is composed of sixteen sequential glottal shapes, were used to perform the quasi-steady and dynamic airflow simulations. The modal displacements of the medial surface of vocal fold were defined by the surface-wave equation proposed by Titze (1984) [5]. For dynamic airflow simulations, the progression from shape to shape was kept continuous, as in normal vocal fold vibration. In addition, all time-dependent terms in the Navier-Stokes equations were included in the numerical solution. For the quasi-steady assumption, each of the sixteen shapes was treated in isolation, without time continuity between the shapes. All time-dependent terms were eliminated from the Navier-Stokes equations to satisfy the quasi-steady conditions. The vibration frequencies of the glottal wall in the dynamic setups were chosen to be 100 Hz and 500 Hz. The setups were implemented with and without a vocal tract for both quasi-steady and dynamic simulations.

## RESULTS AND DISCUSSION

The preliminary results show that adopting the quasi-steady assumption at the low vibration frequency will not cause the simulation results to severely deviate from those of the dynamic situations. However, at the high frequency, the quasi-steady assumption requires significant error correction. Specifically, the peak flow rate, average flow rate, skewness of the flow waveform and wall pressure distribution were compared between the quasi-steady and dynamic simulations, and the differences dramatically increase at the high vibration frequency. Furthermore, including a vocal tract also increases the differences dramatically. A momentum budget analysis was also conducted to compare the unsteady acceleration term and convective acceleration term during the vibration cycle. The unsteady acceleration is found to have the same order of magnitude as the convective acceleration and is non-negligible through the cycle for both frequencies.

## CONCLUSION

The quasi-steady assumption yields acceptable results at the low vibration frequency. Nevertheless, at the high frequency, the quasi-steady assumption seems no longer hold and should be cautiously used.

## ACKNOWLEDGMENTS

This research was funded by the National Institute on Deafness and Other Communication Disorders (NIDCD) grant number 5R21DC016428.

## REFERENCES

[1] Mongeau L, Franchek N, Coker CH, Kubli RA. Characteristics of a pulsating jet through a small modulated orifice, with application to voice production. *J Acoust Soc Am* 1997;102(2):1121-1133.

[2] Zhang Z, Mongeau L, Frankel SH. Experimental verification of the quasi-steady approximation for aerodynamic sound generation by pulsating jets in tubes. *J Acoust Soc Am* 2002;112(4):1652-1663.

[3] Vilain CE, Pelorson X, Fraysse C, Deverge M, Hirschberg A, Willems J. Experimental validation of a quasi-steady theory for the flow through the glottis. *J Sound Vib* 2004;276(3-5):475-490.

[4] Krane MH, Barry M, Wei T. Dynamics of temporal variations in phonatory flow. *J Acoust Soc Am* 2010;128(1):372-383.

[5] Titze IR. Parameterization of the glottal area, glottal flow, and vocal fold contact area. *J Acoust Soc Am* 1984;75(2):570-580.

# AERODYNAMIC ANALYSIS OF THE INSPIRATION FOR A UNILATERAL VOCAL FOLDS PARALYSIS IN A SYNTHETIC LARYNX MODEL

**Marion Semmler[1*], Gregor Peters[1], Anne Schützenberger[1], Andreas Müller[2], Matthias Echternach[3], Stefan Kniesburges[1]**

[1] Div. of Phoniatrics and Pediatric Audiology, Dep. of Otorhinolaryngology, University Hospital Erlangen, Friedrich-Alexander-University Erlangen-Nürnberg

[2] SRH Wald-Klinikum Gera GmbH, Gera – Germany

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

***Keywords*: aerodynamics, inspiration, unilateral paralysis, synthetic mode**

## INTRODUCTION

Unilateral vocal fold (VF) paralysis causes insufficient glottal closure during phonation and partial obstruction of the airways during respiration. The beneficial influence of different therapeutic approaches, i.e., medialization by injection or thyroplasty on the phonation process and the resulting voice quality are well documented. However, the reported effects on the respiration process are ambiguous: spirometric parameters i.e., lung capacity and tidal volume display no significant variation with different degrees of blockage [1] whereas patient questionnaires attest to a subjective impairment of the inspiration, particularly under physical stress [2]. In this study, a systematic analysis of the inspiration process is performed for unilateral VF paralysis.

## METHODS

The 3D-printed synthetic larynx models consisting of a trachea with vocal folds and ventricular folds display variable degrees of obstruction (Fig.1): 25.5° (respiration), 12.75° (intermediate), 8° (paramedian), 0° (median), -5° (compensation). In addition, a vocal tract represents the physiological conditions of the upper airways as a channel of varying cross-sections. In order to simulate the inhalation process at different levels of physical stress (Table 1), a mass flow generator generates constant and temporally varying flow velocities. The resulting pressure and flow resistance is determined for 12 different positions along the synthetic larynx.

## RESULTS

Both at constant and variable flow rates, the pressure and flow resistance increase with rising airflow velocity and growing blockage of the airway, which reflects an amplified respiratory distress. The impairment displays a non-linear behavior with sudden degradation between 0◦/8◦ and between 12.75°/25.5°.

## CONCLUSION

Based on a synthetic larynx model, this study objectively confirms the subjective feeling of inspiration difficulties in patients with unilateral VF paralysis, especially under physical stress. Thus, the planning of phonosurgical interventions should consider the personal circumstances of the patient and strive for a compromise between efficient phonation and reasonable respiration capabilities.

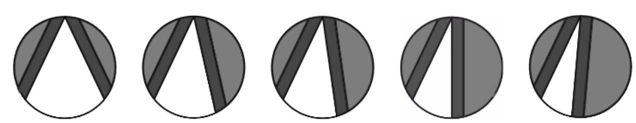



Figure 1. Top-view of 5 different degrees of airway blockage during inspiration process.

**Table 1. Flow velocity at different degrees of physical stress [2].**

| work level | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $v_F$[l/s] | 0.47 | 1.12 | 1.77 | 2.25 | 2.72 | 3.35 |

## REFERENCES

[1] Asik et al. "Airway and respiration parameters improve follow-ing vocal fold medialization: a prospective study." Ann. Otol. Rhinol. Laryngol. 124:972-977, 2015.

[2] Brunner et al, "Subjective breathing impairment in unilateral vocal fold paralysis"," IALP, 63:1-10, 2011.

[3] Naranjo et al, "A nomogram for assessment of breathing patterns during treadmill exercise" Br. J. Sports Med., 35:80-83, 205.

*The 14th International Conference on Advances in Quantitative Laryngology, Voice and Speech Research (AQL)*

# AUTOMATIC SEGMENTATION OF RELATIVE FUNDAMENTAL FREQUENCY FROM CONTINUOUS SPEECH

**Mark Berardi[1], Erin Tippit[2], Gui N. DeSouza[3], Maria Dietrich[1,4]**

[1] Department of Psychiatry and Psychotherapy, University Hospital Bonn, Bonn, Germany
[2] Department of Communication Sciences and Disorders, University of Wisconsin-Madison, Madison, WI, United States
[3] Department of Electrical Engineering and Computer Science, University of Missouri, Columbia, MO, United States
[4] Department of Speech, Language and Hearing Sciences, University of Missouri, Columbia, MO, United States

***Keywords*: Voice; Vocal Fatigue; Vocal Effort; Machine Learning**

## INTRODUCTION

Relative fundamental frequency (RFF) is used as an acoustic correlate for vocal effort [1]. The analysis measures pitch periods of vowels surrounding a voiceless consonant. Typically, this is done through manual processing of the acoustic signal. This approach is limited when applied to a large data set. While there is a tool to provide semi-automated analysis of RFF using a MATLAB program [2], this applies to isolated productions of /afa/, /ifi/, and /ufu/. Therefore, measuring RFF from utterances taken from continuous speech on a large scale is not practical. The purpose of this study was to implement an automatic segmentation pipeline to measure RFF from continuous speech. Here two algorithms were tested, and the reliability and validity were measured to determine the optimal approach.

## METHODS

*Participants*

Ninety-two females from a sEMG study [3] were recorded to assess potential differences in speech acoustics associated with vocal fatigue. Inclusion criteria included ages between 21 and 39 years, native English speaking, typical hearing, and vocally healthy [3]. The participants were recorded with a head-mounted microphone (AKG, Model C520, Vienna, Austria) in a sound-isolation booth (IAC Acoustics, North Aurora, IL, USA) and were sampled at 44.1 kHz with 16-bit quantization.

*Stimuli*

The participants repeated the sentence “The dew shimmered over my shiny blue shell again” 55 times. Then they repeated the sentence “Only we feel you do fail in new fallen dew” 55 times. From these sentences six vowel-consonant-vowel (VCV) speech segments were used for RFF analysis. These were “d**ew shi**mmered,” “m**y shi**ny,” “bl**ue she**ll,” “w**e fee**l,” “d**o fai**l”, and “n**ew fa**llen.”

*Analysis*

An analysis pipeline for RFF measurement was developed as follows. First, the VCV utterances were segmented by aligning the acoustic signals to text using the Penn Phonetics Lab Forced Aligner (PFA) [4]. Two algorithms were compared for automatic RFF calculation. Algorithm 1 was the semi-automated RFF MATLAB algorithm [2] (while this algorithm was not developed for the specific application to RFF speech samples from sentences it is used here for comparison). The proposed Algorithm 2 used a novel approach, also as a MATLAB program. This new MATLAB program used the Hidden Markov Model Speech Recognition Toolkit (HTK) for fricative identification, Praat for pulse detection, and MATLAB to reject unusable samples (criteria for rejection from [5]) and compute the RFF.

The measured RFF values from both algorithms were compared with 20% of the samples analyzed manually. Pearson’s $r$ was used to measure reliability. The root-mean-squared errors (RMSE) of the pulses were also calculated.

## RESULTS

A subset of the audio files was used to validate the broad unaccompanied application of the pipeline. For fricative identification, Algorithm 1 required manual intervention for 4% of the files and Algorithm 2 required manual intervention for 1% of the files.

For Algorithm 1, there was a reliability of $r = 0.86$, and for Algorithm 2, there was a reliability of $r = 0.81$. The RMSE of the calculated RFF are shown in Figure 1.

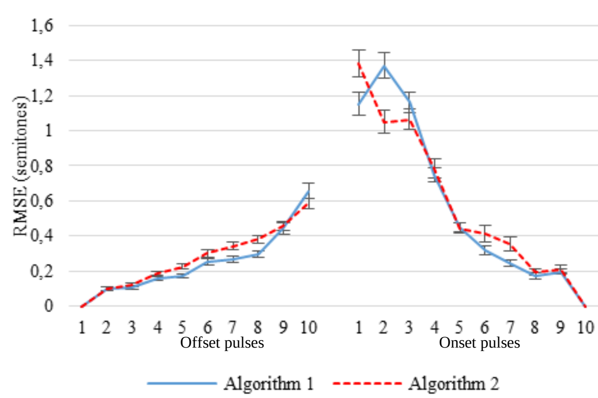



Figure 2. RMSE comparison with standard-error bars of the two algorithms (one using existing RFF calculations and the second using HTK tools and Praat) for each offset and onset cycle.

## DISCUSSION

In terms of reliability and RMSE, the two algorithms performed similarly, with Algorithm 1 performing slightly better. However, Algorithm 2 needed less manual intervention for fricative identification. A combination of the two approaches could provide better results.

## CONCLUSION

Towards scaling RFF measurement of continuous speech, an analysis pipeline with low human intervention was developed and tested. Here the extraction of VCV utterances and measurement of RFF had a low error rate for fricative identification and high reliability with the manual analysis.

## ACKNOWLEDGMENTS

Research funded by NIDCD research grants R15DC015335 and R01DC018026 and “Classifying necksurface EMG signals for the early detection of vocal fatigue in student teachers” and “Neurobiological and psychobiological signatures of vocal effort in early career teachers.” Special thanks to Erin Tippit for manual RFF analyses.

## REFERENCES

[1] Eadie TL, Stepp CE. Acoustic correlate of vocal effort in spasmodic dysphonia. *Ann Otol Rhinol Laryngol*. 2013;122(3):169-176.

[2] Vojtech JM, Segina RK, Buckley DP, Kolin KR, Tardif MC, Noordzij JP, Stepp CE. Refining algorithmic estimation of relative fundamental frequency: Accounting for sample characteristics and fundamental frequency estimation method. *J Acoust Soc Am* 2019;146(5), 3184.

[3] Gao Y, Dietrich M, DeSouza GN. Classification of vocal fatigue using sEMG: Data imbalance, normalization, and the role of Vocal Fatigue Index scores. *Applied Sciences* 2021;11(10):4335.

[4] Jiahong Yuan and Mark Liberman. Speaker identification on the SCOTUS corpus. *J Acoust Soc Am* 2008; 123, 3878-3878.

[5] Lien YS, Stepp CE. "Automated estimation of relative fundamental frequency," *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2013, pp. 2136-2139.

*The 14th International Conference on Advances in Quantitative Laryngology, Voice and Speech Research (AQL)*

# OPENHSV: AN OPEN PLATFORM FOR LARYNGEAL HIGH-SPEED VIDEOENDOSCOPY

**Andreas M Kist[1,2], Stephan Dürr[1], Anne Schützenberger[1], Michael Döllinger[1]**

[1] Div. of Phoniatrics and Pediatric Audiology, University Hospital Erlangen, Friedrich-Alexander-University Erlangen-Nürnberg

[2] Department for Artificial Intelligence in Biomedical Engineering (AIBE), Friedrich-Alexander-University Erlangen-Nürnberg, Germany

***Keywords*: High-Speed Videoendoscopy; Software; Deep Learning; Analysis**

## INTRODUCTION

High-speed videoendoscopy (HSV) is an important tool to study laryngeal dynamics. However, most current commercially available systems are technical outdated and provide only proprietary software and minor image segmentation and quantitative analysis. In this work, we provide a novel open-source system, termed OpenHSV, that is based on state-of-the-art hardware, provides a graphical user interface capable to acquire and analyze data, and that has been evaluated in a clinical context.

## METHODS

The OpenHSV software is written in the scientific programming language Python. The graphical user interface is based on the libraries PyQt5 and pyqtgraph (see Figure 1). Efficient deep neural networks for segmenting the glottal area were setup in TensorFlow 1.15 and trained on the BAGLS dataset [1]. Video and audio data are acquired using a color high-speed camera at 4,000 Hz and a professional lavalier microphone. A preliminary clinical study consisting of 28 healthy subjects was conducted to evaluate the functionality of the OpenHSV platform. To assess image quality, we used the natural image quality evaluator (NIQE) metric [2].

OpenHSV is available at www.anki.xyz/openhsv.

## RESULTS

OpenHSV is an open research platform that can be used by non-specialist personnel. We show that our platform is superior in terms of image quality to existing commercial systems (mean NIQE 13.19 for OpenHSV compared to 28.79 and 22.42 for RGB and monochrome images, respectively). All acquired data were analyzed fully automatically. The computed fundamental frequencies show a strong correlation between glottal area waveform and the audio signal and typically deviate less than 1 Hz. Computed quantitative parameter values are in the expected range for healthy individuals (opening quotient 0.998, mean jitter less than 0.5 ms, high values for CPP and HNR).

## CONCLUSION

With OpenHSV, we provide a state-of-the-art open system for research purposes but also for clinical studies due to its simplicity. It allows the fully automatic, quantitative analysis of synchronously acquired video and audio data. Part of this work is used in a novel clinical tool using latest hardware developments.[1]

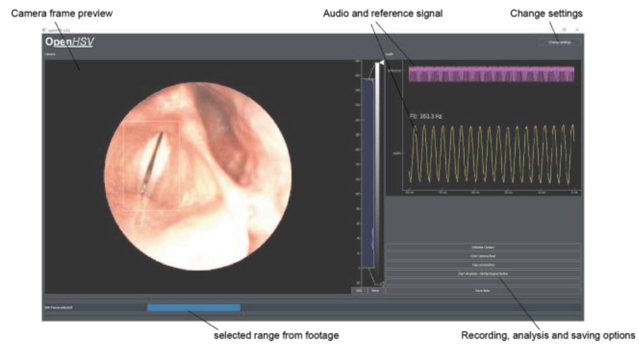



Figure 1. The OpenHSV graphical user interface.

## REFERENCES

[1] Gómez, P., Kist, A. M., Schlegel, P., Berry, D. A., Chhetri, D. K., Dürr, S., ... & Döllinger, M. (2020). BAGLS, a multihospital benchmark for automatic glottis segmentation. Scientific data, 7(1), 1-12.

[2] Mittal, A., Soundararajan, R., & Bovik, A. C. (2012). Making a “completely blind” image quality analyzer. IEEE Signal processing letters, 20(3), 209-212.

[1] https://www.wevosys.com/products/lingwaves4/lingwaves4_high_speed_videoendoscopy.html

# 3D LARYNGEAL IMAGING INTEGRATING PARALLEL OPTICAL COHERENCE TOMOGRAPHY WITH VIDEOSTROBOSCOPY

**Daryush D. Mehta[1, 2], Gopi Maguluri[3], James B. Kobler[1, 2], Nicusor Iftimia[3]**

[1] Massachusetts General Hospital, Boston, MA, USA
[2] Harvard Medical School, Boston, MA, USA
[3] Physical Sciences, Inc., Andover, MA, USA

***Keywords*: 3D laryngeal imaging; phonation; videostroboscopy; vocal folds**

## INTRODUCTION

Currently, laryngologists and speech-language pathologists do not have an endoscopic imaging tool that can directly capture the three-dimensional (3D) surface motion of the vocal folds in real time as patients phonate. The aim of this project is to complement the clinical gold standard of two-dimensional laryngeal videostroboscopy with three-dimensional parallel optical coherence tomography (OCT) for real-time imaging of vocal fold surface and subsurface morphology during phonation. Parallel OCT eliminates motion-blur artifacts exhibited by the sequential sampling of conventional flying-spot OCT due to temporal aliasing when imaging the rapidly moving vocal folds [1, 2].

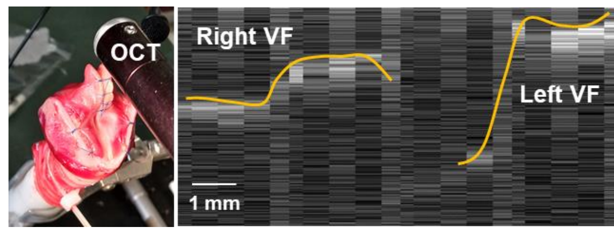
## METHODS

A clinically viable dual channel, transoral endoscopic probe with approximately 10 mm diameter was constructed. Whereas conventional OCT approach scans the laser light across the sample surface, the parallel OCT approach records the interference fringes of multiple illumination points across the sample while the illumination beamlets are stationary, resulting in improved signal-to-noise-ratio. The output power of the OCT swept source is enhanced before splitting into parallel channels for simultaneous interrogation of multiple locations along the vocal folds. A multiple channel digitizer is used to digitize and process the OCT fringes and generate simultaneous OCT reflectivity profiles (A-lines). The recording of the fringes is made only when receiving a trigger signal from a clinical videostroboscopic system. Instrument performance was validated on the bench using aerodynamically driven excised larynx models to simulate voice production.

## RESULTS

The parallel OCT channels recorded cross-sectional images of the vocal fold (similar to single line kymography) to enable the recording of multiple phases of the glottal cycle synchronized with laryngeal videostroboscopy. The strobe triggered flashes at 60 Hz, making the glottal cycle motion appear at a presentation frequency of 0.5 Hz, enabling OCT to gather ~120 temporal phases of the vocal fold cyclic motion in 2 sec. OCT sample illumination consisted of parallel beamlets separated by 420 μm, and thus spanning ~5 mm area. In this preliminary implementation, six OCT beamlets were sequentially moved laterally to reconstruct the full glottal cycle and span 10 mm across the mid-glottis. The full glottal cycle was reconstructed in ~8 sec. The working distance from probe tip was maintained to be 70 mm in accordance with clinical guidance, while achieving a mediolateral sampling resolution of 420 μm over a 10 mm distance and an axial (depth) resolution of 10 μm.

Figure 1 illustrates the excised larynx setup and exemplary parallel OCT image taken during phonation.




DI

Figure 3. Excised larynx setup with parallel optical coherence tomography (OCT) of the right and left vocal fold (VF) during phonation. Surface contours are indicated.

## DISCUSSION & CONCLUSION

Preliminary results are promising to enable spatio-temporal co-registration of two-dimensional color videostroboscopic images and cross-sectional OCT images to ultimately visualize *in vivo* three-dimensional laryngeal imaging in real time. The number of beamlets can be increased using higher-power laser sources to capture more of the vocal fold surface contour during phonation.

## ACKNOWLEDGMENTS

Funding provided by the National Institutes of Health (NIH) National Institute on Deafness and Other Communication Disorders (Grant R43 DC017407). The contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH.

## REFERENCES

[1] Maguluri, G., Mehta, D., Kobler, J., Park, J., and Iftimia, N. (2019). Synchronized, concurrent optical coherence tomography and videostroboscopy for monitoring vocal fold morphology and kinematics. Biomedical Optics Express, 10(9):4450–4461.

[2] Iftimia, N., Maguluri, G., Chang, E., Park, J., Kobler, J., and Mehta, D. (2016). Dynamic vocal fold imaging with combined optical coherence tomography/high-speed video endoscopy. *Proceedings of the 10th International Conference on Voice Physiology and Biomechanics*:1–2.

*The 14th International Conference on Advances in Quantitative Laryngology, Voice and Speech Research (AQL)*

# SPATIAL SEGMENTATION OF HIGH-SPEED VIDEOENDOSCOPY WITH SUB-PIXEL RESOLUTION USING ADAPTIVE THRESHOLDING AND DOUBLE CURVE FITTING

**Hamzeh Ghasemzadeh[1,2,3], David S. Ford[3,4], Maria Powell[5], Dimitar D. Deliyski[3]**

[1] Center for Laryngeal Surgery and Voice Rehabilitation, Massachusetts General Hospital, Boston, MA, USA
[2] Department of Speech, Language & Hearing Sciences, Boston University, Boston, MA, USA
[3] Department of Communicative Sciences and Disorders, Michigan State University, East Lansing, MI, USA
[4] Department of Communication Sciences and Disorders, Thiel College, Greenville, PA, USA
[5] Department of Otolaryngology - Head & Neck Surgery, Vanderbilt University Medical Center, Nashville, TN, USA

***Keywords*: Spatial segmentation, Sub-pixel resolution, High-speed videoendoscopy, Laryngeal imaging**

## INTRODUCTION

High temporal and spatial resolutions of images acquired through high-speed videoendoscopy (HSV) are valuable sources of information for studying different phonatory mechanisms. However, HSV produces a huge amount of data, and their efficient analysis requires the availability of automated processing methods. Of significant importance is the detection of the vocal fold (VF) edges, or the space between them, also known as the glottis. Spatial segmentation is the process that addresses this need. Different spatial segmentation techniques have been proposed previously [1]. The existing approaches may have adequate accuracy and resolution for common analysis based on glottal area waveform. However, recent advances (e.g., spatial calibration [2]) have opened up new avenues for voice science research, which requires a more accurate spatial segmentation method with sub-pixel resolution. This study proposes a new spatial segmentation method that takes full advantage of both temporal and spatial redundancies of HSV frames and hence can achieve sub-pixel resolution.

## METHODS

The proposed method is constructed based on three main rules. (1) The two VFs cannot pass each other. (2) Two adjacent points on a VF are part of the same connected tissue and hence cannot move independently from each other (i.e., spatial redundancy). (3) Positions of a point on a VF in consecutive frames are not independent of each other (i.e., temporal redundancy).

Kymograms of HSV data were created by the algorithm at contiguous scanning lines. An adaptive thresholding technique was devised for estimating the coarse location of each VF edge at each line scan, and a spline curve was fitted on each VF edge estimation. This step exploits the high temporal redundancy of the HSV data and significantly improves the segmentation outcome. Additionally, it ensures locations of the VF edges are continuous in time. Since each kymogram was segmented independently, the spatial redundancy of the data has not been used. Therefore, in the next phase, two spline curves (one per each VF) were fitted on the outcome of segmentations from all kymograms at each specific frame. This step ensures the locations of the VF edges are continuous in space.

To evaluate the performance of the method, ground truths with sub-pixel resolutions were required. To that end, a manual segmentation software was developed that allowed piecewise linear segmentation of the VF edges. This approach can provide an analytic description for very complex contours if a high number of edges are used. Three experts were trained with this software and were tasked to segment 110 frames (100 consecutive frames with 10% random redundancy appended to the end) from different HSV recordings. Regions with high discrepancies between the experts were determined and subsequently presented to all experts for making proper adjustments and reconciliation. Spline curves were used to fuse the outcomes from different experts and also to exploit the temporal and spatial redundancies of the data and to create the ground truth.

## RESULTS AND CONCLUSION

Table 1 presents the average intersection over union (IoU) scores for the initial and reconciliation phases of manual segmentation and the automated method for three different HSV files. Based on table 1, the proposed automated method has comparable performance with expert-labeled data.

**Table 1. IoU scores of manual and automated methods**

| Video ID | Initial | Reconciliation | Automated |
|---|---|---|---|
| **7** | 0.67 | 0.8 | 0.82 |
| **8** | 0.6 | 0.65 | 0.68 |
| **17** | 0.8 | 0.82 | 0.75 |

## REFERENCES

[1] Andrade-Miranda, G., Stylianou, Y., Deliyski, D. D., Godino-Llorente, J. I., & Henrich Bernardoni, N. (2020). Laryngeal image processing of vocal folds motion. Applied Sciences, 10(5), 1556.

[2] Ghasemzadeh, H., Deliyski, D. D., Hillman, R. E., & Mehta, D. D. (2021). Method for Horizontal Calibration of Laser-Projection Transnasal Fiberoptic High-Speed Videoendoscopy. Applied Sciences, 11(2), 822.

*The 14th International Conference on Advances in Quantitative Laryngology, Voice and Speech Research (AQL)*

# ACOUSTIC IDENTIFICATION OF THE VOICING BOUNDARY DURING INTERVOCALIC OFFSETS AND ONSETS BASED ON VOCAL FOLD VIBRATORY MEASURES

**Jennifer M. Vojtech[1,2*], Dante D. Cilento[2], Austin T. Luong[2], Jacob P. Noordzij, Jr.[2], Manuel Diaz-Cadiz[2], Matti D. Groll[1,2], Daniel P. Buckley[2,3], J. Pieter Noordzij[3], Cara E. Stepp[1,2,3]**

[1] Department of Biomedical Engineering, Boston University, Boston, MA 02215, USA
[2] Department of Speech, Language, and Hearing Sciences, Boston University, Boston, MA 02215, USA
[3] Department of Otolaryngology—Head and Neck Surgery, Boston University School of Medicine, Boston, MA 02118, USA
[*]Current Affiliation: Delsys, Inc. and Altec, Inc., Natick, MA 01760, USA; jvojtech@delsys.com

***Keywords*: Voice; Laryngeal Tension; High-Speed Videoendoscopy; Relative Fundamental Frequency**

## INTRODUCTION

Current methods for automating estimates of relative fundamental frequency (RFF), an acoustic indicator of laryngeal tension, rely on manual unvoiced/voiced (U/V) boundary detection from acoustic signals. The aim of this work was to determine the potential benefit of incorporating features derived from true vocal fold vibratory transitions—as characterized using high-speed videoendoscopy—for acoustic U/V boundary detection for automated RFF estimation. It was hypothesized that incorporating features related vocal fold vibratory offsets and onsets would improve acoustic U/V boundary detection accuracy over methods that did not leverage these tuned features.

## METHODS

### *Participants*

Adults with typical voices (N=69) or with a voice disorder characterized by excessive laryngeal tension (N=53) were enrolled in the study. Participants with a voice disorder were either diagnosed with idiopathic Parkinson's disease by a neurologist (25/53) or with a hyperfunctional voice disorder by a board-certified laryngologist (28/53), including: muscle tension dysphonia (20/28), vocal fold nodules (4/28), vocal fold polyp (2/28), vocal fold scarring (1/28), and hyperdermal lesion with secondary supraglottic compression (1/28).

### *Instrumentation and Measurement*

Simultaneous acoustic and high-speed videoendoscopic recordings were collected as 122 participants produced the voiced–unvoiced–voiced utterance, /ifi/. Participants produced /ifi/ utterances at different vocal rates and levels of vocal effort to alter the stiffness of the laryngeal musculature to, in turn, produce voice with varying degrees of laryngeal muscle tension [1].

### *Analysis*

Kinematic time points were extracted from the /ifi/ productions to mark the physiological termination or initiation of vocal fold vibration. A stepwise binary logistic regression was performed to identify acoustic features that coincided with these vocal fold vibratory transitions. A recent version of the RFF algorithm ("aRFF-AP" [2]) was updated with the resulting, physiologically tuned acoustic features to create "aRFF-APH." Chi-square tests were performed to compare U/V boundary detection accuracy relative to the ground-truth videoendoscopic signal between aRFF-APH, aRFF-AP, and manual RFF estimation (i.e., the current gold-standard technique for calculating RFF).

## RESULTS

U/V boundary detection accuracy significantly differed by RFF estimation method for voicing offsets and onsets. Of 7721 productions, 76.0% of boundaries were accurately identified via the aRFF-APH algorithm, compared to 70.3% with the aRFF-AP algorithm and 20.4% with manual estimation.

## DISCUSSION AND CONCLUSIONS

These findings demonstrate that using physiologically tuned acoustic features—which corresponded with the offset and onset of vocal fold vibration—led to significant improvements in U/V boundary detection accuracy that surpassed the gold-standard method for calculating RFF.

## ACKNOWLEDGMENTS

This work was supported by NSF Grant #1247312 (J.M.V.) and NIH Grant #DC015570 (C.E.S.).

## REFERENCES

[1] Stepp CE, Hillman RE, Heaton JT. A virtual trajectory model predicts differences in vocal fold kinematics in individuals with vocal hyperfunction. *J Acoust Soc Am 2010*; 127:3166–3176.

[2] Vojtech JM, et al. Refining algorithmic estimation of relative fundamental frequency: Accounting for sample characteristics and fundamental frequency estimation method. *J Acoust Soc Am* 2019; 146:3184–3202.

# IN VIVO MEASUREMENT OF INTRAGLOTTAL, SUBGLOTTAL, AND VOCAL FOLD COLLISION PRESSURES IN A HEMILARYNGECTOMY PATIENT

**Daryush D. Mehta[1,2,3], James B. Kobler[1,3], Steven M. Zeitels[1,3], Robert H. Petrillo[1], Robert E. Hillman[1,2,3]**

[1] Massachusetts General Hospital, Boston, MA, USA
[2] MGH Institute of Health Professions, Charlestown, MA, USA
[3] Harvard Medical School, Boston, MA, USA

***Keywords*: vocal fold collision; intraglottal pressure; subglottal pressure; voice disorders**

## INTRODUCTION

The direct measurement of vocal fold collision pressure is challenging to carry out *in vivo*, and only two published studies have attempted to gather data from sensors placed intraglottally during phonation [1, 2]. Expectations with respect to the waveform shape of the intraglottal pressure signal come from numerical models of phonation, self-oscillating physical models of synthetic vocal fold–like material, aerodynamically driven excised larynx models, and in vivo animal work. Taken together, this body of literature provides strong evidence that the *in vivo* intraglottal pressure signal during phonation should have two primary components: (1) an impulsive peak in the direction of increasing pressure at the start of the phonatory closed phase (collision/impact component), which is followed in time by (2) a more rounded peak during the phonatory open phase (acoustic pressure component).

The purpose of this presentation is to report on the first *in vivo* application of a recently developed transoral dual-sensor pressure probe that measures intraglottal, subglottal, and vocal fold collision pressures during phonation [3].

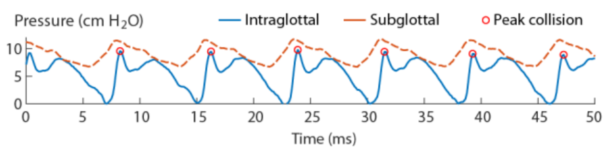
## METHODS

Simultaneous measurement of intraglottal and subglottal pressures was accomplished using two miniature pressure sensors mounted on the end of a transoral cannula and positioned intraglottally in a 78-year-old male who had previously undergone a hemilaryngectomy to treat laryngeal cancer. The patient produced a sustained vowel as the endoscopist stabilized the custom probe against the nonvibrating vocal fold and synchronously recorded laryngeal high-speed videoendoscopy.

## RESULTS

Endoscopic visualization of the larynx using high-speed videoendoscopy enabled positioning of the dual-sensor pressure probe such that the proximal sensor was positioned intraglottally and the distal sensor subglottally. Sustained phonation was captured at 81.4 dB SPL re

**Fig. 4: Intraglottal and subglottal pressure signals measured by the custom probe with peak collision pressures marked.**




15 cm and 126.1 Hz fundamental frequency. Figure 1 displays the intraglottal and subglottal pressure waveforms during phonation exhibiting the two expected components due to vocal fold collision and open phase intraglottal pressure. The ratio of the mean peak collision pressure (9.0 cm $H_2O$) and subglottal pressure (9.0 cm $H_2O$) during the vowel was in line with *in vivo* and excised larynx data in the literature.

## DISCUSSION & CONCLUSION

The results successfully demonstrate feasibility of *in vivo* measurement of vocal fold collision pressures in individuals with a unilateral cordectomy, motivating ongoing data collection that is designed to aid in the development of vocal dose measures that incorporate vocal fold impact stress/collision.

## ACKNOWLEDGMENTS

Funding provided by the Voice Health Institute and the National Institutes of Health (NIH Grant P50 DC015446 awarded to REH). The paper's contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH. Disclosure: Drs. Hillman, Zeitels, Kobler, and Mehta have a financial interest in InnoVoyce LLC, a company focused on developing and commercializing technologies for the prevention, diagnosis, and treatment of voice-related disorders.

## REFERENCES

[1] Gunter *et al.* (2005). Journal of Speech, Language, and Hearing Research, vol. 48, pp. 567–576.

[2] Verdolini *et al.* (1999). Journal of Voice, vol. 13, pp. 184–202.

[3] Mehta *et al.* (2019). Applied Sciences, vol. 9, p. 4360.

# INDIRECT SPATIAL CALIBRATION OF THE HORIZONTAL PLANE OF ENDOSCOPIC LARYNGEAL IMAGES: HOW TO DO IT AND WHAT TO LOOK FOR

**Hamzeh Ghasemzadeh[1,2], Dimitar D. Deliyski[3], Robert E. Hillman[1,4], Daryush D. Mehta[1,4]**

[1] Center for Laryngeal Surgery and Voice Rehabilitation, Massachusetts General Hospital, Boston, MA, USA
[2] Department of Speech, Language & Hearing Sciences, Boston University, Boston, MA, USA
[3] Department of Communicative Sciences and Disorders, Michigan State University, East Lansing, MI, USA
[4] Department of Surgery, Harvard Medical School, Boston, MA, USA

***Keywords*: Horizontal calibrated measurements, Calibrated intraoperative image, Laryngeal imaging**

## INTRODUCTION

Calibrated horizontal-plane measurements from laryngeal images could contribute significantly to refining evidence-based practice and developing patient-specific models and precision-medicine approaches. Laser-projection systems can address the need for direct calibrated measures and provide calibrated horizontal and vertical measurements with sub-millimeter accuracies [1,2]. However, these systems are not widely and commercially available which could pose significant challenges for applications requiring calibrated spatial measurements. This study presents the framework for an alternative approach. The general idea of this indirect approach is to use a common object as a scale for the normalization of the region of interest (ROI). The proposed framework derives the required conditions for validity of outcomes from this indirect method.

## METHODS

The pixel size can provide a convenient way for estimating the *mm*-length of an object from its pixel-length computed from an image. The proposed framework is based on a mathematical analysis of the pixel size. Assuming an imaging system where pixel size only depends on the vertical distance [3], three main conditions are derived. The first condition is the *registration accuracy*, which assumes that the common attribute can be registered accurately across different images. The second condition is the *size consistency of the common attribute*, which assumes that the *mm*-length of the common attribute does not change across different images. Finally, the last condition is the *similarity in vertical distance between the ROI and the common attribute*, which assumes that the ROI and the common attribute are on the same horizontal plane. Additionally, two data-driven tests were developed for evaluating the first and the second conditions, whereas the effect of violation of the third condition was quantified using optical principles and a mathematical model for image formation. Application of the developed framework and the proposed tests were demonstrated using a pre-existing dataset, selecting the vocal fold as the ROI, and four different selections of the common attributes.

## RESULTS AND CONCLUSION

Table 1 compares the investigated common attributes in terms of the framework's three conditions. In conclusion, the vocal fold width offered the best trade-off and seems to be the most proper common attribute for indirect calibration of laryngeal images.

## ACKNOWLEDGMENTS

Funding provided by the Voice Health Institute and the National Institutes of Health (NIH Grants R01 DC017923, R01 DC007640, P50 DC015446). The paper's contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH. Disclosure: Drs. Robert Hillman and Daryush Mehta have a financial interest in InnoVoyce LLC, a company focused on developing and commercializing technologies for the prevention, diagnosis, and treatment of voice-related disorders.

## REFERENCES

[1] Ghasemzadeh, H., Deliyski, D. D., Hillman, R. E., & Mehta, D. D. (2021). Method for Horizontal Calibration of Laser-Projection Transnasal Fiberoptic High-Speed Videoendoscopy. Applied Sciences, 11(2), 822.

[2] Ghasemzadeh, H., Deliyski, D. D., Ford, D. S., Kobler, J. B., Hillman, R. E., & Mehta, D. D. (2020). Method for vertical calibration of laser-projection transnasal fiberoptic high-speed videoendoscopy. Journal of Voice, 34(6), 847-861.

[3] Ghasemzadeh, H., & Deliyski, D. D. (2020). Non-Linear Image Distortions in Flexible Fiberoptic Endoscopes and their Effects on Calibrated Horizontal Measurements Using High-Speed Videoendoscopy. Journal of Voice, In press.

**Table 1. Comparing suitability of different common attributes for indirect calibration of vocal folds.**

| Common attribute | Registration accuracy | Size consistency | Similarity in the vertical distance |
|---|---|---|---|
| **Vocal fold length** | Highest | Lowest | High |
| **Vocal fold width** | High | High | High |
| **Blood vessel on vocal fold** | Lowest | Low | High |
| **Blood vessel on a nearby tissue** | Low | Highest | Low |

# EFFECTS OF VERTICAL GLOTTAL DUCT LENGTH ON INTRAGLOTTAL PRESSURES IN THE UNIFORM AND CONVERGENT GLOTTIS

**Sheng Li[1], Ronald C. Scherer[2]**

[1] College of Science, Xijing University, 1 Xijing Road, Xi'an 710123, People's Republic of China

[2] Department of Communication Sciences and Disorders, Bowling Green State University, Bowling Green, Ohio 43403, USA

***Keywords*: Vertical Glottal Duct Length; Glottal Angle; Intraglottal Pressure; Vocal Fold Geometry**

## INTRODUCTION

Glottal geometry parameters play an important role during the process of phonation, which include glottal angle [1], glottal inferior and superior vocal fold surfaces [2], glottal entrance and exit radii [3, 4], and the symmetry and obliquity of the glottis [5]. Besides these parameters, the vertical glottal duct length, which is described as "the axial length of the glottis or the upstream-to-downstream length of the glottis" [6, 7], should also have significant effects on the intraglottal aerodynamic parameters, such as intraglottal air pressure distributions and velocity profiles as well as the bulk glottal flow and glottal airflow resistance, in addition to influencing the aerodynamic forces acting upon the vocal folds. The vertical length of the glottis is also a determining factor of phonatory threshold pressure (PTP) [8].

The purpose of this study was to explore the effects of the vertical glottal duct length on the intraglottal aerodynamic parameters and the entrance and transglottal pressure coefficients, and the associated potential effects on the PTP for uniform and convergent glottal angles. If the intraglottal air pressure forces are highly sensitive to different vertical glottal duct lengths, this critical phase of the vocal fold motion may vary significantly with duct length. In addition, since higher pitches are associated with longer vocal folds and thus shorter vertical glottal duct lengths, understanding the aerodynamic effects of the vertical glottal duct length change is a basic aspect to understanding phonation and fundamental frequency.

## METHODS

*Experimental dimensions*

The vertical glottal duct length is defined here as the length between the glottal entrance and the glottal exit (Figure 1). In order to cover the possible range of the vertical glottal duct length during phonation, four values were chosen, 0.108, 0.308, 0.608, and 0.908 cm. One uniform glottal angle (for the uniform glottis), and three typical convergent glottal full angles (-5°, -10° and -20° (for the convergent glottis)), three transglottal pressures (500, 1000, and 1500 Pa), and three typical minimal glottal diameters (0.01, 0.04, 0.16 cm) were selected and

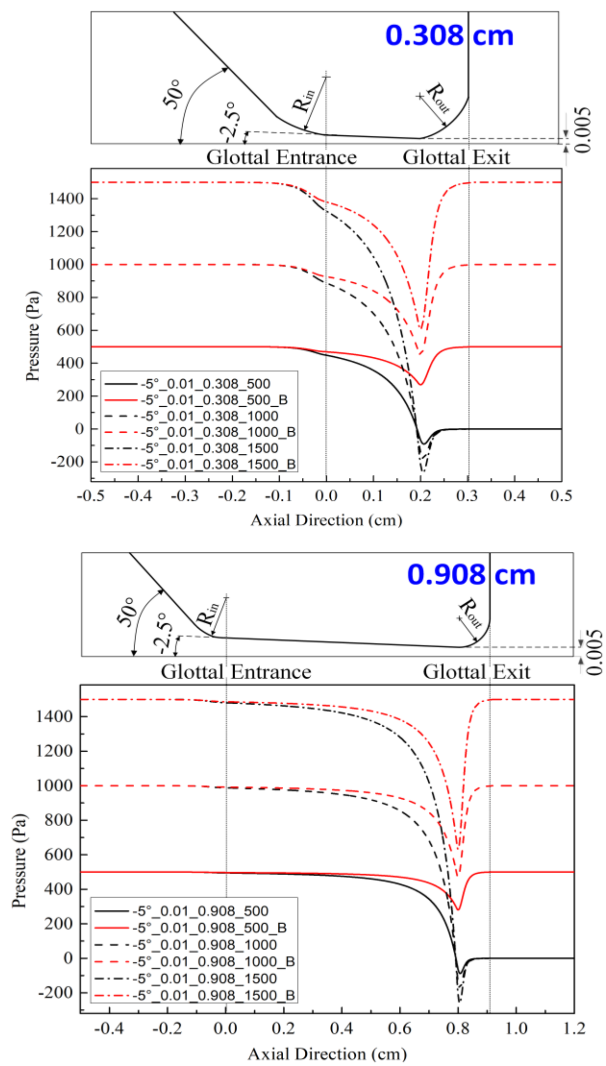



**Fig. 4: The outline of the glottal configuration for 0.308 and 0.908 cm vertical (axial) glottal duct length and 0.01 cm minimal glottal diameter for a −5° (convergent) glottal angle. Distances are in centimeters. The lower figures show the pressure distributions for 3 transglottal pressure values (black) and lossless Bernoulli (red).**

used for each of the four vertical glottal duct lengths. Parameters that were held constant were the glottal entrance radius $R_{in}$ (0.15 cm), the glottal exit radius $R_{out}$ (0.108 cm), the inferior vocal fold surface angle (50° relative to the tracheal wall), the superior vocal fold surface angle (90° relative to the vertical direction), the upstream (tracheal) inlet computational domain length (0.2 cm), the downstream outlet computational domain length (0.5 cm), and the anterior-posterior glottal length (1.2 cm). The glottis was two-dimensional (rectangular in the anterior-posterior direction or transverse plane).

*Computational Method*

ANSYS Fluent (ANSYS, Inc., Canonsburg, PA; http://www.ansys.com/Products/Fluids/ANSYS-Fluent) was employed to obtain the pressures and flows. The code solved the Navier-Stokes equations for laminar and incompressible airflow. Grids were generated by Gambit (ANSYS, Inc., Canonsburg, PA), with both structured and unstructured meshes, with 412,000 to 1,720,000 nodes, made finer in regions where pressure was expected to change quickly. The flow field was assumed to be symmetric across the midline of the glottis; only the half flow field was modeled.

## RESULTS

The vertical glottal duct length has significant effects on transglottal pressure and intraglottal pressure distributions for both the uniform and convergent glottis. For the uniform glottis: (1) A longer vertical glottal duct length increases the intraglottal and transglottal pressures for a constant flow, and more so for smaller glottal diameters. (2) The transglottal pressure coefficient is significantly increased by the added flow resistance of a longer duct length (range: 1.1 – 108). (3) The value of the PTP expression is greatly affected by duct length increase as both the duct length and the transglottal pressure coefficient are in the expression. (4) The glottal entrance pressure coefficient is highly dependent on the vertical glottal duct length only for lower flows and Reynolds numbers, and is relatively independent of duct length, glottal diameter, and transglottal pressure above a flow value of approximately 50 $cm^3$/s; the entrance pressure coefficient is a relatively local phenomenon for flows higher than about 50 $cm^3$/s.

For the convergent glottis: (1) A longer vertical glottal duct length increases the intraglottal pressures (Fig. 1), causing greater aerodynamic forces on the vocal fold medial surfaces, and more so for smaller glottal minimal diameters. (2) The glottal entrance loss coefficient typically increases with vertical glottal duct length. (3) For the smallest diameter, the entrance loss coefficient was the largest; the coefficient value decreased as the Reynolds number or convergent glottal angle increased. (4) The transglottal pressure coefficient somewhat increased as the vertical glottal duct length increased, especially for low flows and small glottal minimal diameters. (5) The transglottal pressure coefficient was largest for the smallest diameter (range: 1.1 – 2.2), and decreased as the diameter or Reynolds number increased. (6) The vertical glottal duct length has only small effects on the exit coefficient and volume flow. (7) A longer vertical glottal duct length results in a less abrupt increase of the intraglottal flow velocity and wall shear stress.

## CONCLUSION

This study suggests that a longer vertical glottal duct length (a) increases the intraglottal pressures, and more so for larger transglottal pressures, and (b) significantly increases the transglottal pressure coefficient, especially for low flows and small glottal minimal diameters, for both uniform and convergent glottal shapes. Also, the entrance pressure coefficient is a relatively local phenomenon which decreases for longer ducts for the convergent glottis. These results suggest that the vertical glottal duct length plays an important role in voice production, and should be well specified when building computational and physical models of the vocal folds.

## ACKNOWLEDGMENTS

Research was funded by the National Natural Science Foundation of China (NSFC, No. 11974289) and the Key Research and Development Program of Shaanxi Province, China (No. 2019SF-103).

## REFERENCES

[1] Li S, Scherer RC, Wan M, Wang S, Wu H. The effect of glottal angle on intraglottal pressure. *J Acoust Soc Am* 2006;119(1):539-548.

[2] Li S, Scherer RC, Wan M, Wang S, Wu H. Numerical study of the effects of inferior and superior vocal fold surface angles on vocal fold pressure distributions. *J Acoust Soc Am* 2006;119(5):3003-3010.

[3] Scherer RC, De Witt KJ, Kucinschi BR. The effect of exit radii on intraglottal pressure distributions in the convergent glottis. *J Acoust Soc Am* 2001;110(5):2267-2269.

[4] Li S, Scherer RC, MingXi W, SuPin W. The effect of entrance radii on intraglottal pressure distributions in the divergent glottis. *J Acoust Soc Am* 2012;131(2):1371-1377.

[5] Scherer RC, Shinwari D, De Witt KJ, Zhang C, Kucinschi BR, Afjeh AA. Intraglottal pressure profiles for a symmetric and oblique glottis with a divergence angle of 10 degrees. *J Acoust Soc Am* 2001;109(4):1616-1630.

[6] Li S, Scherer RC, Fulcher LP, Wang X, Qiu L, Wan M, et al. Effects of Vertical Glottal Duct Length on Intraglottal Pressures and Phonation Threshold Pressure in the Uniform Glottis. *J Voice* 2018;32(1):8-22.

[7] Li S, Scherer RC, Wan M. Effects of Vertical Glottal Duct Length on Intraglottal Pressures in the Convergent Glottis. *Applied Sciences* 2021;11(10):1-25.

[8] Titze IR. The physics of small-amplitude oscillation of the vocal folds. *J Acoust Soc Am* 1988;83(4):1536-1552.

# LIST OF AUTHORS

| AUTHORS | EMAIL | AFFILIATION |
| --- | --- | --- |
| Abberton Evelyn | e.abberton@ucl.ac.uk | University College London, UK |
| Abur Defne | dabur@bu.edu | Boston University, USA |
| Acevedo Karol | karol.acevedo.encalada@gmail.com | Department of Communication Sciences and Disorder, Universidad San Sebastian, |
| Adkins Lacey | Lacey.Adkins@fmolhs.org | Dept. of Otolaryngology-Head and Neck Surgery, Louisiana State University Health Sciences Center, New Orleans, USA. Voice Center, The Our Lady of The Lake Regional Medical Center, Baton Rouge, USA |
| Aichinger Philipp | philipp.aichinger@meduniwien.ac.at | Department of Otorhinolaryngology, Division of Phoniatrics-Logopedics, Medical University of Vienna, Austria |
| Alzamendi Gabriel Alejandro | galzamendi@ingenieria.uner.edu.ar | Institute for Research and Development on Bioengineering and Bioinformatics, CONICET-UNER, Entre Rios, Argentina |
| Anderson David | anderson@gatech.edu | |
| Angerer Susanne | | Div. of Phoniatrics and Pediatric Audiology, Dept. of Otorhinolaryngology, University Hospital Erlangen, Germany |
| Araneda Oscar | | Department of Communication Sciences and Disorders, Universidad de los Andes, |
| Atará-Piraquive Ángela Patricia | apatarap@unal.edu.co | Department of collective Health, Universidad Nacional de Colombia |
| Avhad Amit | amit.g.avhad@vanderbilt.edu | Department of Mechanical Engineering, Vanderbilt University, Nashville, TN, USA |
| Banks Russell | rbanks@mghihp.edu | |
| Berardi Mark | Mark.Berardi@ukbonn.de | Department of Psychiatry and Psychotherapy, University Hospital Bonn, Bonn, |
| Berry David A | daberry@ucla.edu | Department of Head and Neck Surgery, David Geffen School of Medicine at UCLA, Los Angeles, California, USA |
| Bhaskaran Divya | bhas0021@umn.edu | |
| Böhringer David | david.boehringer@fau.de | Dept. of Physics, Biophysics Group, FAU Erlangen-Nürnberg, Erlangen, Germany |
| Bottalico Pasquale | pb81@illinois.edu | |
| Buckley Daniel | buckleyd@bu.edu | Department of Speech, Language, and Hearing Sciences, Boston University, Boston, MA 02215, USA. Department of Otolaryngology—Head and Neck Surgery, Boston University School of Medicine, Boston, MA 02118, USA |
| Bulusu Sridhar | sridharbulusuvie@gmail.com | Medical University of Vienna, Department of Otorhinolaryngology, Division of Phoniatrics-Logopedics, Vienna, Austria |
| Burns James | jburns0@partners.org | |
| Calvache Mora Carlos Alberto | cabbeto@gmail.com | Department of Mechatronics Engineering, Universidad Militar Nueva Granada, Bogotá, Colombia. Vocology Center, Bogotá, Colombia |
| Cantor-Cutiva Lady Catherine | lccantorc@unal.edu.co | Universidad Nacional de Colombia - Universidad Manuela Beltrán |
| Canziani Herbert | herbert.canziani@fau.de | Dept. of Chemical and Biological Engineering, Chair of Particle Technology, FAU Erlangen-Nürnberg, Erlangen, Germany |
| Cárdenas Jenny | jfcardenasma@unal.edu.co | Universidad Nacional de Colombia, Bogotá, Colombia |
| Castillo-Allendes Adrian | casti208@msu.edu | Department of Communicative Sciences and Disorders, Michigan State University, East Lansing, Michigan, USA |
| Castro Christian | christian.castro@uv.cl | Department of Communication Sciences and Disorders, Universidad de los Andes, |
| Charney Sara | charney.sara@mayo.edu | |
| Chen Mo | | |
| Chhetri Dinesh | dchhetri@mednet.ucla.edu | Department of Head & Neck Surgery, UCLA, Los Angeles, California, United |
| Ciccarelli Gregory | | Massachusetts Institute of Technology, USA |
| Cilento Dante | | Department of Speech, Language, and Hearing Sciences, Boston University, Boston, MA 02215, USA |
| Cortés Juan Pablo | jpcortes@mgh.harvard.edu | Center for Laryngeal Surgery and Voice Rehabilitation Laboratory, Massachusetts General Hospital, MA, United States |
| Dahl Kimberly | dahl@bu.edu | Boston University, USA |
| Dalmasso Maria Del Carmen | fonomdalmasso@gmail.com | |
| De Alarcón Alessandro | | |
| Delgado Mauricio | mauricio.delgado@uc.cl | Departamento de Ciencias de la Salud, Facultad de Medicina, Pontificia Universidad Católica de Chile, Santiago, Chile |
| Deliyski Dimitar | ddd@msu.edu | Department of Communicative Sciences and Disorders, Michigan State University, East Lansing, MI, USA |
| Demolin Didier | didier.demolin@sorbonne-nouvelle.fr | |
| Desouza Guilherme | DeSouzaG@missouri.edu | Department of Electrical Engineering and Computer Science, University of Missouri, Columbia, MO, United States |
| Devaraj Vinod | n11935890@students.meduniwien.ac.at | Department of Otorhinolaryngology, Division of Phoniatrics-Logopedics, Medical University of Vienna, Austria<br>Department of Signal Processing and Speech Communication Laboratory, Technische Universität Graz, Austria |
| Diaz-Cadiz Manuel | mdiazcad@bu.edu | Department of Speech, Language, and Hearing Sciences, Boston University, Boston, MA 02215, USA |
| Dietrich Maria | maria.dietrich@ukbonn.de | Department of Psychiatry and Psychotherapy, University Hospital Bonn, Bonn, Germany<br>Department of Speech, Language and Hearing Sciences, University of Missouri, |
| Dinther Joost J.s. | | European Institute for ORL-HNS, Otorhinolaryngology & Head and Neck Surgery, GZA Sint-Augustinus, Belgium |
| Dolling Anton | adolling@bu.edu | |
| Döllinger Michael | Michael.Doellinger@uk-erlangen.de | Division of Phoniatrics and Pediatric Audiology, Department of Otorhinolaryngology, University Hospital Erlangen, FAU Erlangen-Nürnberg, |
| Dürr Stephan | stephan.duerr@uk-erlangen.de | Div. of Phoniatrics and Pediatric Audiology, University Hospital Erlangen, Friedrich-Alexander-University Erlangen-Nürnberg |

| AUTHORS | EMAIL | AFFILIATION |
|---|---|---|
| Echternach Matthias | matthias.echternach@med.uni-muenchen.de | Div. of Phoniatrics and Pediatric Audiology, Dept. of Otorhinolaryngology, Munich University Hospital (LMU), Germany |
| Elangovan Naveen | naveen@umn.edu | |
| Erath Byron D. | berath@clarkson.edu | Department of Mechanical and Aeronautical Engineering, Clarkson University, Potsdam, NY, USA |
| Espinoza Victor | vespinoza@uchile.cl | Department of Sound, Universidad de Chile, Santiago, Chile |
| Falk Sebastian | sebastian.falk@uk-erlangen.de | Div. of Phoniatrics and Pediatric Audiology, Dept. of Otorhinolaryngology, University Hospital Erlangen, Germany |
| Farbos De Luzan Charles | farboscs@ucmail.uc.edu | |
| Feng Chuayo | | |
| Ford David S. | forddav5@msu.edu | Department of Communicative Sciences and Disorders, Michigan State University, East Lansing, MI, USA<br>Department of Communication Sciences and Disorders, Thiel College, Greenville, |
| Fourcin Adrian | a.fourcin@ucl.ac.uk | Laryngograph Ltd, Wallington UK, Emeritus Professor at University College |
| Fox Annie | afox-galalis@mghihp.edu | |
| Gao Yixiang | yg5d6@mail.missouri.edu | Department of Electrical Engineering and Computer Science, University of Missouri, Columbia, MO, United States |
| Geng Biao | biao.geng@maine.edu | Department of Mechanical Engineering, University of Maine, Orono, Maine, USA |
| Gerstenberger Claus | | Division of Phoniatrics, ENT University Hospital Graz, Medical University Graz, |
| Ghasemzadeh Hamzeh | ghasemza@msu.edu | Center for Laryngeal Surgery and Voice Rehabilitation, Massachusetts General Hospital, Boston, MA, USA<br>Department of Speech, Language & Hearing Sciences, Boston University, Boston, MA, USA<br>Department of Communicative Sciences and Disorders, Michigan State University, |
| Gostian Antoniu-Oreste | | Dept. of Otorhinolaryngology, University Hospital Erlangen, FAU Erlangen- |
| Grillo Elizabeth | egrillo@wcupa.edu | Department of Communication Sciences and Disorders, West Chester University, West Chester, PA, USA |
| Groll Matti | mgroll@bu.edu | Department of Biomedical Engineering, Boston University, Boston, MA, USA<br>Department of Speech, Language, and Hearing Sciences, Boston University, |
| Gugatschka Markus | markus.gugatschka@medunigraz.at | Division of Phoniatrics, ENT University Hospital Graz, Medical University Graz, |
| Gutmark Ephraim | gutmarej@ucmail.uc.edu | |
| Guzman Marco | guzmann.marcoa@gmail.com | Department of Communication Sciences and Disorders, Universidad de los Andes, |
| Hablani Surbhi | shablani@bu.edu | Department of Speech, Language, and Hearing Sciences, Boston University, |
| Hesse Nicolas | | Dept. of Chemical and Biological Engineering, Chair of Particle Technology, FAU Erlangen-Nürnberg, Erlangen, Germany |
| Hillman Robert E | hillman.robert@mgh.harvard.edu | Center for Laryngeal Surgery and Voice Rehabilitation, Massachusetts General Hospital, Boston, MA, U.S.A.<br>MGH Institute of Health Professions, Massachusetts General Hospital, Boston, MA, U.S.A. |
| Howard Ian | ian.howard@plymouth.ac.uk | SECAM, University of Plymouth, Plymouth, UK |
| Hron Tiffiny | tainsworth@mgh.harvard.edu | |
| Hunter Eric J. | ejhunter@msu.edu | Department of Communicative Sciences and Disorders, Michigan State University, East Lansing, Michigan, USA |
| Ibarra Emiro | emiro.ibarra@sansano.usm.cl | Department of Electronic Engineering, Federico Santa María Technical University, Valparaíso, Chile |
| Iftimia Nicusor | iftimia@psicorp.com | Physical Sciences, Inc., Andover, MA, USA |
| Ikuma Takeshi | tikuma@lsuhsc.edu | Dept. of Otolaryngology-Head and Neck Surgery, Louisiana State University Health Sciences Center, New Orleans, USA<br>Voice Center, The Our Lady of The Lake Regional Medical Center, Baton Rouge, |
| Inostroza-Moreno Gonzalo | gonzalo.inostroza.moreno@gmail.com | |
| Ip Carman | agirlcalledcarman@gmail.com | Department of Chinese and Bilingual Studies, The hong Kong Polytechnic University, Hong Kong |
| Ishikawa Keiko | ishikak@illinois.edu | |
| Jakubaß Bernhard | bernhard.jakubass@uk-erlangen.de | Division of Phoniatrics and Pediatric Audiology, Department of Otorhinolaryngology, University Hospital Erlangen, FAU Erlangen-Nürnberg, |
| Jerez Rodrigo Maximiliano | rodrigomjerez@yahoo.com.ar | Facultad de Ciencias de la Salud, Universidad Católica de La Plata, La Plata, Buenos Aires, Argentina |
| Jr. Jacob | | Department of Speech, Language, and Hearing Sciences, Boston University, Boston, MA 02215, USA |
| Kaltenbacher Manfred | manfred.kaltenbacher@tugraz.at | Inst. of Fundamentals and Theory in Electrical Eng., Div. Vibro- and Aeroacoustics, Graz University of Technology, Austria |
| Kapsner-Smith Mara | mkapsner@uw.edu | |
| Khosla Siddarth | khoslasm@uc.edu | |
| Kimberley Teresa | tkimberley@mghihp.edu | MGH Institute of Health Profesions, USA |
| Kist Andreas M. | andreas.kist@uk-erlangen.de | Div. of Phoniatrics and Pediatric Audiology, Dept. of Otorhinolaryngology, University Hospital Erlangen, Germany |
| Kniesburges Stefan | stefan.kniesburges@uk-erlangen.de | Div. of Phoniatrics and Pediatric Audiology, Dept. of Otorhinolaryngology, University Hospital Erlangen, Germany |
| Kobler James | James.Kobler@mgh.harvard.edu | Massachusetts General Hospital, Boston, MA, USA<br>Harvard Medical School, Boston, MA, USA |
| Konczak Jurgen | jkonczak@umn.edu | |
| Krusemark Carol | ckrusemark@partners.org | |
| Kumar Pravin S. | pravinkumars@ssn.edu.in | Sri Sivasubramaniya Nadar College of Engineering, Department of Biomedical Engineering, Kalavakkam, Chennai, India |

| AUTHORS | EMAIL | AFFILIATION |
| --- | --- | --- |
| Kunduk Melda | mkunduk@lsu.edu | Dept. of Otolaryngology-Head and Neck Surgery, Louisiana State University Health Sciences Center, New Orleans, USA<br>Voice Center, The Our Lady of The Lake Regional Medical Center, Baton Rouge, USA |
| Kuo Yi-Ling | kuoy@upstate.edu | Upstate Medical University, USA |
| Kwong Elaine | speech@polyu.edu.hk | Department of Chinese and Bilingual Studies, The hong Kong Polytechnic University, Hong Kong |
| Leite Danilo Rangel Arruda | danilorangel@buscapb.com.br | Departamento de Estatística. Programa de Pós-graduação em Modelos de Decisão em Saúde. Universidade Federal da Paraíba, João Pessoa, Paraíba, Brasil<br>Empresa Brasileira de Serviços Hospitalares- Ebserh, João Pessoa, Paraíba, Brasil |
| Li Sheng | sheng@mail.xjtu.edu.cn | |
| Lopes Leonardo Wanderley | lwlopes@hotmail.com | Departamento de Estatística. Programa de Pós-graduação em Modelos de Decisão em Saúde. Universidade Federal da Paraíba, João Pessoa, Paraíba, Brasil<br>Departamento de Fonoaudiologia. Universidade Federal da Paraíba, João Pessoa, |
| Luo Haoxiang | haoxiang.luo@vanderbilt.edu | Department of Mechanical Engineering, Vanderbilt University, Nashville, TN, USA |
| Luong Austin, Noordzij | | Department of Speech, Language, and Hearing Sciences, Boston University, Boston, MA 02215, USA |
| Müller Andreas | | SRH Wald-Klinikum Gera GmbH, Gera – Germany |
| Maguluri Gopi | gmaguluri@psicorp.com | Div. of Phoniatrics and Pediatric Audiology, University Hospital Erlangen, Friedrich-Alexander-University Erlangen-Nürnberg<br>Department for Artificial Intelligence in Biomedical Engineering (AIBE), Friedrich-Alexander-University Erlangen-Nürnberg, Germany |
| Mahnan Arash | amahnan@umn.edu | |
| Manriquez Rodrigo | rodrigo.manriquezp@usm.cl | Department of Electronic Engineering, Universidad Técnica Federico Santa María, Valparaíso, Chile |
| Marchetta Jeffrey | | |
| Marks Katherine L | katherine.l.marks@gmail.com | Center for Laryngeal Surgery and Voice Rehabilitation, Massachusetts General Hospital, Boston, MA, U.S.A.<br>MGH Institute of Health Professions, Massachusetts General Hospital, Boston, MA, |
| Maryn Youri | youri.maryn@gza.be | European Institute for ORL-HNS, Otorhinolaryngology & Head and Neck Surgery, GZA Sint-Augustinus, Belgium |
| Maurerlehner Paul | paul.maurerlehner@tugraz.at | Inst. of Fundamentals and Theory in Electrical Eng., Div. Vibro- and Aeroacoustics, Graz University of Technology, Austria |
| Mcglashan Julian | Julian.McGlashan@nottingham.ac.uk | ENT Department, Nottingham University Hospitals, Nottingham, UK |
| Mcwhorter Andrew | feedback@marybird.com | Dept. of Otolaryngology-Head and Neck Surgery, Louisiana State University Health Sciences Center, New Orleans, USAVoice Center, The Our Lady of The Lake Regional Medical Center, Baton Rouge, USA |
| Mehta Daryush D | mehta.daryush@mgh.harvard.edu | MGH Institute of Health Professions, Charlestown, MA, USA<br>Massachusetts General Hospital, Boston, MA, USA<br>Harvard Medical School, Boston, MA, USA |
| Mendes Clarion | cmendes2@illionis.edu | |
| Moran Camilo | | Department of Communication Sciences and Disorders, Universidad de los Andes, |
| De Moraes Ronei Marcos | ronei@de.ufpb.br | Departamento de Estatística. Programa de Pós-graduação em Modelos de Decisão em Saúde. Universidade Federal da Paraíba, João Pessoa, Paraíba, Brasil |
| Motie-Shirazi Mohsen | | Department of Mechanical and Aeronautical Engineering, Clarkson University, Potsdam, NY, USA |
| Movahhedi Mohammadreza | mohammadreza.movahhedi@maine.edu | Department of Mechanical Engineering, University of Maine, Orono, Maine, USA |
| Muise Jason | jmuise1@partners.org | |
| Müller Sarina K. | | Dept. of Otorhinolaryngology, University Hospital Erlangen, FAU Erlangen- |
| Murgia Silvia | smurgia2@illinois.edu | University of Illinois at Urbana-Champaign, USA |
| Naghibolhosseini Maryam | naghib@msu.edu | Department of Communicative Sciences and Disorders, Michigan State University, East Lansing, Michigan, USA |
| Noordzij J Pieter | Pieter.Noordzij@bmc.org | Department of Otolaryngology—Head and Neck Surgery, Boston University School of Medicine, Boston, MA 02118, USA |
| Nudelman Charles | cnudelman@mghihp.edu | MGH Institute of Health Professions, USA |
| Oh Jinseok | ohxxx414@umn.edu | University of Minnesota, USA |
| Omotara Gbenga | goowfd@mail.missouri.edu | University of Missouri, USA |
| Orbelo Diana | orbelo.diana@mayo.edu | Clínica Mayo, USA |
| Oren Liran | orenl@ucmail.uc.edu | |
| Orlikoff Robert F | | College of Allied Health Sciences, East Carolina University, Greenville, North |
| Ortiz Andrew J | aortiz15@mgh.harvard.edu | Center for Laryngeal Surgery and Voice Rehabilitation, Massachusetts General Hospital, Boston, MA, U.S.A |
| Palaparthi Anil | anil.palaparthi@utah.edu | The National Center for Voice and Speech, University of Utah, Salt Lake City, |
| Parra Jesús | jesus.parrap@sansano.usm.cl | Department of Electronic Engineering, Federico Santa María Technical University, Valparaíso, Chile |
| Peñuela Lina | lina.penuela@unimilitar.edu.co | Department of Mechatronics Engineering, Universidad Militar Nueva Granada, |
| Peters Gregor | gregor.peters@uk-erlangen.de | Div. of Phoniatrics and Pediatric Audiology, Dept. of Otorhinolaryngology, University Hospital Erlangen, Germany |
| Peterson Sean D. | peterson@uwaterloo.ca | Mechanical and Mechatronics Engineering Department, University of Waterloo, Waterloo, ON, N2L 3G1 , Canada |
| Petrillo Robert | | Massachusetts General Hospital, Boston, MA, USA |
| Peukert Wolfgang | | Dept. of Chemical and Biological Engineering, Chair of Particle Technology, FAU Erlangen-Nürnberg, Erlangen, Germany |

| AUTHORS | EMAIL | AFFILIATION |
| --- | --- | --- |
| Powell Maria | maria.e.powell@vanderbilt.edu | Department of Otolaryngology - Head & Neck Surgery, Vanderbilt University Medical Center, Nashville, TN, USA |
| Prado Pavel | | Latin American Brain Health Institute (BrainLat), Universidad Adolfo Ibáñez, Santiago de Chile, Chile |
| Quatieri Thomas F | quatieri@ll.mit.edu | Massachusets Institute of Technology, USA |
| Quezada Camilo | | Department of Communication Sciences and Disorders, Universidad de Chile, |
| Redman Yvonne | ygredman@illinois.edu | |
| Rousseau Bernard | rousseau.csd@pitt.edu | Department of Communication Science and Disorders, University of Pittsburgh, Pittsburgh, PA, USA |
| Roy Nelson | nelson.roy@health.utah.edu | University of Utah, Salt Lake City, UT, USA |
| Sayce Lea | lea.sayce@pitt.edu | Department of Communication Science and Disorders, University of Pittsburgh, Pittsburgh, PA, USA |
| Scherer Ronald C. | ronalds@bgsu.edu | |
| Schlegel Patrick | pschlegel@ucla.edu | Department of Head & Neck Surgery, UCLA, Los Angeles, California, United |
| Schoder Stefan | stefan.schoder@tugraz.at | |
| Schützenberger Anne | anne.schuetzenberer@uk-erlangen.de | Div. of Phoniatrics and Pediatric Audiology, Dept. of Otorhinolaryngology, University Hospital Erlangen, Germany |
| Segina Roxanne | rsegina@bu.edu | Department of Speech, Language, and Hearing Sciences, Boston University, |
| Semmler Marion | marion.semmler@uk-erlangen.de | Div. of Phoniatrics and Pediatric Audiology, Dep. of Otorhinolaryngology, University Hospital Erlangen, Friedrich-Alexander-University Erlangen-Nürnberg |
| Serry Mohamed | mserry@uwaterloo.ca | Mechanical and Mechatronics Engineering Department, University of Waterloo, Waterloo, ON, N2L 3G1, Canada |
| Sogg Stephanie | ssogg@mgh.harvard.edu | Massachusetts General Hospital, Boston, MA, USA Harvard Medical School, Boston, MA, USA |
| Solaque Leonardo | leonardo.solaque@unimilitar.edu.co | Department of Mechatronics Engineering, Universidad Militar Nueva Granada, |
| Stadelman-Cohen Tara | tstadelmancohen@mgh.harvard.edu | MGH Institute of Health Profesions, USA |
| Stepp Cara | cstepp@bu.edu | Department of Biomedical Engineering, Boston University, Boston, MA 02215, USA Department of Speech, Language, and Hearing Sciences, Boston University, Boston, MA 02215, USA Department of Otolaryngology—Head and Neck Surgery, Boston University School |
| Stewart Molly | stewarme@clarkson.edu | Department of Mechanical and Aeronautical Engineering, Clarkson University, Potsdam, NY, USA |
| Stinnett Sandra | | |
| Stipancic Kaila | klstip@buffalo.edu | University at Buffalo, Buffalo, NY, USA |
| Subaciute Austeja | asub@bu.edu | |
| Sundström Elias | sundstes@uc.edu | Cincinnati College of Medicine, USA |
| Svec Jan | jan.svec@upol.cz | Palacky University Olomouc, the Czech Republic |
| Sweet Bridget | bsweet@illinois.edu | |
| Testart Alba | | Department of Communication Disorders, Universidad de Playa Ancha, Valparaíso, |
| Tippit Erin, | | Department of Communication Sciences and Disorders, University of Wisconsin-Madison, Madison, WI, United States |
| Titze Ingo | ingo.titze@utah.edu | The National Center for Voice and Speech, University of Utah, Salt Lake City, |
| Toles Laura E | ltoles@mghihp.edu | MGH Institute of Health Professions, Charlestown, MA, USA Massachusetts General Hospital, Boston, MA, USA |
| Tracy Lauren | Lauren.Tracy@bmc.org | Department of Otolaryngology-Head and Neck Surgery, Boston University School of Medicine, Boston, MA |
| Van Leer Eva | evanleer@gsu.edu | Georgia State University, USA |
| Van Mersbergen Miriam | Miriam.van.Mersbergen@Memphis.edu | The University of Memphis, USA |
| Van Stan Jarrad | jvanstan@mghihp.edu | MGH Institute of Health Professions, USA |
| Velasco Alexandra | alexandra.velasco@unimilitar.edu.co | Department of Mechatronics Engineering, Universidad Militar Nueva Granada, |
| Veltrup Reinhard | reinhard.veltrup@uk-erlangen.de | Div. of Phoniatrics and Pediatric Audiology, Dept. of Otorhinolaryngology, University Hospital Erlangen, Germany |
| Verdi Alessandra | averdi@mghihp.edu | MGH Institute of Health Professions, Boston, MA, USA Massachusetts General Hospital, Boston, MA, USA |
| Verguts Monique | | European Institute for ORL-HNS, Otorhinolaryngology & Head and Neck Surgery, GZA Sint-Augustinus, Belgium |
| Vojtech Jennifer | jmvojtech@gmail.com | Department of Biomedical Engineering, Boston University, Boston, MA, USA Department of Speech, Language, and Hearing Sciences, Boston University, |
| Wang Xiaojian | xiaojian.wang@maine.edu | Department of Mechanical Engineering, University of Maine, Orono, Maine, USA |
| Watson Peter | pjwatson@umn.edu | |
| Weerathunge Hasini | asiniw@bu.edu | Department of Biomedical Engineering, Boston University, Boston, MA Department of Speech, Language, and Hearing Sciences, Boston University, |
| Wendler Olaf | | Div. of Phoniatrics and Pediatric Audiology, Dept. of Otorhinolaryngology, University Hospital Erlangen, Germany |
| Wigley Sarah | swjohn@illinois.edu | |
| Wilson Azure | | Department of Communication Science and Disorders, University of Pittsburgh, Pittsburgh, PA, USA |
| Wurzinger Andreas | | |
| Xue Qian | qian.xue@maine.edu | Department of Mechanical Engineering, University of Maine, Orono, Maine, USA |

| AUTHORS | EMAIL | AFFILIATION |
|---|---|---|
| Yousef Ahmed M | yousefah@msu.edu | Department of Communicative Sciences and Disorders, Michigan State University, East Lansing, Michigan, USA |
| Zacharias Stephanie R. C. | Zacharias.Stephanie@mayo.edu | Mayo Clinic-Arizona, Scottsdale, Arizona, USA<br>Department of Otolaryngology-Head and Neck Surgery, Mayo Clinic, Phoenix, |
| Zañartu Matías | matias.zanartu@usm.cl | Department of Electronic Engineering, Federico Santa María Technical University, Valparaíso, Chile |
| Zeitels Steven | zeitels.steven@mgh.harvard.edu | Massachusetts General Hospital, Boston, MA, USA<br>MGH Institute of Health Professions, Charlestown, MA, USA<br>Harvard Medical School, Boston, MA, USA |
| Zheng Xudong | xudong.zheng@maine.edu | Department of Mechanical Engineering, University of Maine, Orono, Maine, USA |
| Zilker Julian | julian.zilker@gmx.de | Div. of Phoniatrics and Pediatric Audiology, Dept. of Otorhinolaryngology, University Hospital Erlangen, Germany |
| Zillig Tobias | tobias.zillig@fau.de | Div. of Phoniatrics and Pediatric Audiology, Dept. of Otorhinolaryngology, University Hospital Erlangen, Germany |

AQL 2021