



**16TH INTERNATIONAL CONFERENCE ADVANCES IN
QUANTITATIVE LARYNGOLOGY, VOICE, AND SPEECH
RESEARCH (AQL)**

BOOK OF ABSTRACTS

JUNE 23-27, 2025
GRONINGEN, THE NETHERLANDS



TABLE OF CONTENTS

Podium Session 1.....	1
Vocal Fold Vibratory Kinematics in Pressed and Breathy Phonation.....	1
Test-Retest Reliability of the Daily Phonotrauma Index: Normal Variance Across Short- and Long-Term Time Scales.....	3
Towards Reproducible Acoustic Measurements Obtained During Rigid Laryngoscopy with an Endoscope-Mounted Microphone.....	5
Podium Session 2.....	7
Tailored Stimuli to Investigate the Neural Coding of Speech.....	7
Sensorimotor Control of Vocal Pitch Across Age in Adulthood.....	9
Voice Loudness, Self-Assessment and Auditory-Sensory Feedback in Voice Production in Parkinson's Disease.....	11
Podium Session 3.....	13
Deep Learning-Based Laryngo-Tracheal Landmark Tracking for Rapid and Safe Vocal Fold Injection.....	13
Mechanisms of Vocal Fold Inferior Medial Bulging with Thyroarytenoid Muscle Activation.....	15
Engineered Composite Hydrogels for In Situ Tissue Repair: Enhanced Adhesion, Porosity, And Immunomodulation.....	17
Psychosocial Stress-Induced Changes in Vocal Fold Epithelial Barrier Integrity.....	19
Podium Session 4.....	21
Injectable Click Decellularized Extracellular Matrix Hydrogels for Vocal Fold Tissue Engineering: An In Vivo Biocompatibility Study.....	21
A Multi-Material Hydrogel Model to Study Chronic Fibrosis in Vocal Folds.....	23
Innovating Treatments for Vocal Fold Scarring: Controlled Dexamethasone Delivery with a Novel Light-Activatable Implant.....	25
Poster Session 1.....	27
Effects Of Surgery on The Relationship Between Subglottic Pressure and Fundamental Frequency in Vocal Fold Dynamics in Patients with Benign Laryngeal Diseases.....	27
Improving Vocal Fold Paralysis Detection from Voice Recordings Using a Robust ML Mode.....	29
Use Cases for Optopalatographic Articulatory Measurements: A Portable and Easy to Use Solution.....	31
Clustering Study of Voice Characteristics of Mandarin Speakers in Lifespan.....	33
Enhancing Glottis Segmentation in Laryngeal High-Speed Videoendoscopy Using Yolo-Based ROI Detection.....	35
Visual Perceptual Analysis of Vibratory Patterns and Mucosal Wave Differences Between Vocally Healthy Children and Children with Vocal Fold Nodules.....	37
Feedback And Feedforward Auditory-Motor Processes for Voice and Articulation in Hearing Loss.....	39
A Dosimetry Comparative Study in Dear Evan Hansen.....	41
Voice Cepstral Peak Prominence Metrics in Italian by Age, Sex, And Pre-Processing Method.....	43
Exploring The Use of Amniotic Materials to Prevent Vocal Fold Scar After Injury.....	45
Histological Evaluation of Human Laryngeal Embryogenesis.....	46
Laryngoresponders: Are They More Common in Individuals Diagnosed with Vocal Hyperfunction?	48
Post-Influenza Airway Injury: Immune Cell Recruitment and Oncogene Myc-Driven Basal Cell Regeneration.....	50
Comparability And Quantification of Vocal Tract-Related Efficiency Strategies of Different Voice Qualities in Four Professional Singers.....	51
Vowel Onset VRT: A Comparison of Signal Envelope Algorithms and Their Key Parameters.....	53
Pitch Reflex Responses in Individuals with Hearing Impairment Via Acoustics and Electroencephalography.....	55
Role Of Piezo1 in Human Vocal Fold Fibroblasts.....	57
Podium Session 5.....	59
An Update on Conversation Training Therapy (CTT)	59
The Effect of Listener Feedback on Sensorimotor Adaptation in Articulation and Prosody.....	61
Effects Of Delayed Auditory Feedback on Speech Kinematics.....	63
Sibilant Differentiation Following Tongue Cancer Surgery: Evidence from Longitudinal Articulatory-Kinematics 6- And 12- Months Post-Surgery.....	65
Podium Session 6.....	67
Spatiotemporal Indices of Relative Fundamental Frequency for Quantifying Intraspeaker Variability in Vocal Hyperfunction.....	67
Differences In Ambulatory Sympathetic Arousal Between Patients with Vocal Hyperfunction and Vocally Healthy Controls.....	69
Effects Of Cognitive Load On Vocal Motor Control: Changes in Subglottal Pressure and Laryngeal Muscle Activation....	71
Podium Session 7.....	73
Transverse Ultrasound Imaging of the Phonating Larynx.....	73

Asynchronous Parallel Reinforcement Learning for Voice Control in a Human Larynx Model.....	75
Quantifying The Vocal Fold Oscillations Using Dynamic 3D Magnetic Resonance Imaging.....	77
Poster Session 2.....	79
Quantifying Anatomic, Morphologic, Acoustic, and Perceptual Characteristics of Tracheoesophageal Voice Pre and Post Balloon Dilation Utilizing High Speed Digital Imaging.....	79
Measuring The Effects of Corticosteroid Use on Vocal Fold Inflammation and Biomechanics Using a Rabbit Model.....	81
Development and Validation of Personal Technology to Treat Hypophonia in Patients with Parkinsons Disease.....	82
Modeling Uncertainty in Speech Adaptation to Sensory Feedback Perturbations.....	84
Identifying the Speech Breathing Kinematics of Older Adults with Presbyphonia.....	86
Differences in Vocal Fold Vibratory Kinematics Between Women with Phonotrauma and Vocally Healthy Controls.....	88
Comparison of Salivary Cortisol Levels Across Laryngeal Pathologies.....	90
Targeted Nebulized ECM Therapy for Vocal Fold Healing.....	92
Discovering The Effects of Vascular Lesions of the Vocal Fold Via Computational Modeling.....	94
A Platform for Monitoring Ecological Vocal Efficiency.....	95
Dynamic 3D Vocal Fold MRI For Quantification of Vocal Fold Oscillations in Different Voice Production Mechanisms.....	97
Electroglottographic Analysis of the Voice in Mandarin-Speaking Children.....	99
A Tool for Efficient HSV Rating Based on the Voice-Vibratory Assessment with Laryngeal Imaging (VALI) Form.....	101
Temporal Changes in Relative Fundamental Frequency in Continuous Speech: Contrasting Effects of State Vocal Fatigue and Vocal Fatigue Index Scores.....	102
Classification of Adductor Spasmodic Dysphonia with Audio and Glottal Area Waveform Fusion.....	104
Speech Intelligibility in Speakers with Adductor Laryngeal Dystonia (ADLD).....	106
The Effect of Intentional Voice Quality Alteration on Mental Effort Captured Via Pupillometry and Via The NASA-TLX And Borg CR-10 Scales.....	118
A Total Variation Based Semi-Automatic High-Speed Video Denoising Pipeline.....	110
Podium Session 8.....	112
Arytenoid Adduction and Medialization: Resolving Height Asymmetry and Posterior Gap to Optimize Vocal Efficiency..	112
A Microengineered Vocal Fold Model for High-Through Screening Applications.....	114
Simulation Of Flow Induced Vibration of the Vocal Folds Using the Lattice Boltzmann Method.....	116
A Feasibility Study of Subharmonic Voice Detection by Deep Learning.....	118
Podium Session 9.....	120
Acoustic-Aerodynamic Vocal Efficiency Metrics Identify Changes in Vocal Function Following Voice Therapy for Phonotraumatic Vocal Hyperfunction Subgroups.....	120
Respiratory Kinematics in Aging Singers and Non-Singers: Relationship to Vocal Effort.....	122
Modeling Laryngeal Motor Control with Somatosensory Feedback for Prephonatory Vocal Posture.....	124

PODIUM SESSION 1

Vocal Fold Vibratory Kinematics in Pressed and Breathy Phonation

Laura E. Toles¹, Avery Moore^{1,2}, Ted Mau¹

¹ Department of Otolaryngology – Head & Neck Surgery, University of Texas Southwestern Medical Center, Dallas, TX, USA

² Department of Speech, Language, and Hearing, University of Texas at Dallas, Richardson, TX, USA

Keywords: High speed videoendoscopy; Pressed phonation; Breathy phonation

Abstract

Introduction & rationale: Two common types of phonation that a healthy mechanism can produce include pressed and breathy phonation. Pressed phonation is associated with low glottal airflow, increased perceived effort to produce sound, and increased vocal fold adduction (Patel et al., 2022), but it may not result in an aberrant vocal quality to an untrained listener. Alternately, breathy phonation is characterized by increased glottal airflow and incomplete glottal closure. The increased airflow seen in breathy phonation auditorily results in turbulent noise in the voice (Fukazawa et al., 1988). Although a healthy vocal mechanism can produce each of these types of phonation, they can potentially be a precursor for a vocal injury or imply a voice disorder is already present, particularly if used exclusively or regularly (Hillenbrand & Houde, 1996; Hillman et al., 1989).

Phonotraumatic vocal hyperfunction is associated with the formation of benign vocal fold lesions (e.g., vocal fold polyps, nodules) due to increased impact stress and inflammation of the vocal folds (Hillman et al., 2020). The etiology and behaviors associated with phonotrauma can be attributed to a combination of factors, including increased vocal intensity (e.g., persistent loud voicing), heightened vocal dosage (e.g., heavy voice use), and hyperadduction of the vocal folds or increased laryngeal forces during phonation (Kunduk & McWhorter, 2009). Since pressed phonation is assumed to be associated with increased adduction of the vocal folds, persistently using pressed phonation can theoretically contribute to the development of phonotraumatic lesions, especially in the setting of heavy voice use (Toles et al., 2021). Relatedly, breathy phonation can also signal the presence of a variety of voice disorders, such as glottal insufficiency (e.g., vocal fold immobility), a lesion on the vocal folds impeding complete glottal closure, or non-phonotraumatic vocal hyperfunction (e.g., muscle tension dysphonia). The physiological underpinnings of these phonation types requires more study, as literature correlating non-modal phonation with vocal fold kinematics is sparse. One study has demonstrated promise for the use of high-speed videoendoscopy (**HSV**) to quantify the kinematics of different types of phonation, such as pressed and breathy phonation (Patel et al., 2022). That study showed preliminary evidence that peak closing velocity varied across phonation types as well as a trend of vibratory amplitude decreasing from breathy to neutral to pressed phonation. However, the study was limited by a small sample size. Better understanding of the impact of non-modal phonation at the vocal fold level can help clinicians detect patients who exhibit inefficient vocal function that may indicate a current voice disorder or increase the risk of developing a disorder over time.

Objectives: The primary objective of this study was to identify differences in vocal fold vibratory kinematics between typical, pressed, and breathy phonation conditions and to determine whether differences varied based on sex.

Methods: Forty vocally healthy adults (32 female; 8 male) participated in the study. Participants reported no current or chronic voice problems and presented with a typical larynx during screening laryngoscopy. All participants underwent HSV via a transoral rigid scope while phonating on a sustained /i/ at a comfortable speaking pitch and volume in three phonatory modes: typical, pressed, and breathy. Video exams were recorded at 4000 frames per second at 512 x 512 resolution and were processed using the Glottis Analysis Tools software [version 2020] (Kist et al., 2021). Analyses were completed on 200 consecutive vibratory cycles. Vibratory kinematic parameters were chosen for analysis based on their physiological connection to opening and/or closing phases of the vibratory cycle. As such, symmetry measures were not analyzed for the purpose of this study. HSV parameters (Open Quotient, Stiffness Index, Closing Quotient, Amplitude-to-Length Ratio, Speed Index, and Peak Closing Velocity) were each compared across phonation types using a mixed analysis of variance (ANOVA). Phonation type was entered as a within-subjects variable, and sex was included as a between-subjects variable. Tukey's corrected post hoc comparisons were conducted for significant interaction effects or phonation type main effects.

Results: There were no significant interaction effects between phonation type and sex across all ANOVA models. All models except for Speed Index had significant phonation type main effects, all of which can be visualized in **Figure 1**. Post hoc comparisons revealed that pressed phonation had lower Open Quotient ($p < .001$), higher Stiffness Index ($p < .001$), lower Closing Quotient ($p < .001$), and lower Peak Closing Velocity ($p < .001$) compared to typical phonation. Pressed phonation had lower amplitude-to-length ratio than breathy phonation ($p = .048$) but did not differ from typical phonation ($p = .068$). Typical and breathy phonation only significantly differed on Closing Quotient ($p = .004$), which showed that breathy phonation had higher values.

There was a significant main effect of sex for Open Quotient ($p = .028$) and Closing Quotient ($p = .031$), both of which had higher values in females.

Conclusions: Differences in phonation type were evident in vocal fold kinematic parameters. Most of the differences were seen between pressed phonation and typical phonation, whereas breathy phonation was generally similar to typical phonation. The kinematic measures suggest that pressed phonation results in increased adduction reducing the percentage of time the vocal folds are open, increased speed with which the glottis opens and closes, and longer closing time than opening time. There were differences between sexes, but sex did not interact with phonation type. These findings introduce distinct patterns to differentiate pressed from typical phonation which can be helpful when identifying aberrant vibratory patterns that might not be obvious to an untrained listener.

References:

- Fukazawa, T., El-Assuoaty, A., & Honjo, I. (1988). A new index for evaluation of the turbulent noise in pathological voice. *The Journal of the Acoustical Society of America*, 83(3), 1189-1193.
- Hillenbrand, J., & Houde, R. A. (1996). Acoustic correlates of breathy vocal quality: Dysphonic voices and continuous speech. *Journal of Speech and Hearing Research*, 39(2), 311-321. <Go to ISI>://A1996UE60100007
- Hillman, R. E., Holmberg, E. B., Perkell, J. S., Walsh, M., & Vaughan, C. (1989). Objective assessment of vocal hyperfunction: An experimental framework and initial results. *Journal of Speech and Hearing Research*, 32(2), 373-392.
- 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.
- Kist, A. M., Gómez, P., Dubrovskiy, D., Schlegel, P., Kunduk, M., Echternach, M., Patel, R., Semmler, M., Bohr, C., & Dürr, S. (2021). A deep learning enhanced novel software tool for laryngeal dynamics analysis. *Journal of Speech, Language, and Hearing Research*, 64(6), 1889-1903.
- Kunduk, M., & McWhorter, A. J. (2009). True vocal fold nodules: The role of differential diagnosis. *Current Opinion in Otolaryngology & Head & Neck Surgery*, 17(6), 449-452.
- Patel, R. R., Sundberg, J., Gill, B., & Lå, F. M. (2022). Glottal airflow and glottal area waveform characteristics of flow phonation in untrained vocally healthy adults. *Journal of Voice*, 36(1), 140. e141-140. e121.
- Toles, L. E., Ortiz, A. J., Marks, K. L., Burns, J. A., Hron, T., Van Stan, J. H., Mehta, D. D., & Hillman, R. E. (2021). 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*, 30(1), 199-209.

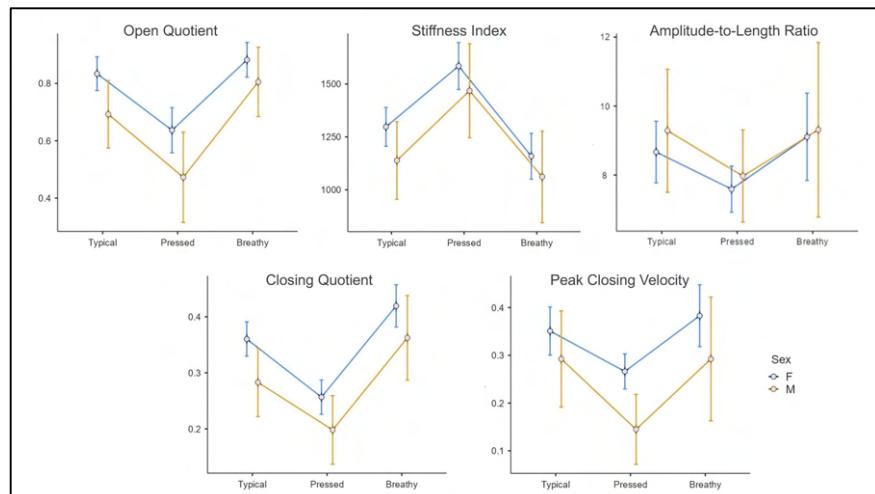


Figure 1. Estimated marginal means and 95% confidence intervals of kinematic parameters that demonstrated a significant phonation type main effect. Females are presented in blue and males in orange.

Test-retest reliability of the Daily Phonotrauma Index: Normal variance across short- and long-term time scales

Jameson Cooper¹, Benjamin Kevelson¹, Hamzeh Ghasemzadeh^{1,3}, Daryush Mehta^{1,2,3}, Robert Hillman^{1,2,3}, Jarrad Van Stan^{1,2,3}

¹ Center for Laryngeal Surgery and Voice Rehabilitation, Massachusetts General Hospital, Boston, MA, USA

² MGH Institute of Health Professions, Boston, MA, USA

³ Department of Surgery, Harvard Medical School, Boston, MA

Keywords: ambulatory voice monitoring, vocal hyperfunction, vocal fold nodules

Introduction & rationale:

Phonotraumatic Vocal Hyperfunction (PVH) is a class of voice disorders characterized by chronic trauma to the vocal folds, often resulting in vocal fold nodules or polyps. Ambulatory voice monitoring allows for the analysis of real-world vocal behaviors in patients with voice disorders. The Daily Phonotrauma Index (DPI) combines two features from ambulatory voice monitoring to represent the potential for phonotrauma (Van Stan et al, 2020b), differentiate patients from controls (Van Stan et al, 2020a), track change following treatment (Van Stan et al, 2020b, 2021), and correlate with self-reported vocal status changes in daily life (Nudelman et al, 2022). However, to date, there has been no study of the normal variance or test-retest reliability of this measure. Quantifying this measure, especially with respect to phonation type (e.g., speech versus singing), can aid with future work in interpreting clinically-meaningful change for patients with PVH.

Objectives:

This study aims to evaluate the test-retest reliability of the DPI in vocally-healthy speakers across short-term (within a single week) and long-term (across 6 months) time scales in addition to studying the normal variance of singing versus speech as they relate to the DPI.

Methods:

125 vocally-healthy females—matched in age, sex, and occupation to patients with PVH—wore an ambulatory voice monitor for approximately seven days. Summary statistics for the DPI and its contributing features (sound pressure level [SPL] skewness, H1-H2 standard deviation) were calculated for each day. To address short-term reliability, within-week daily recordings were randomly assigned to two groups consisting of three days each, a time scale that previous work has shown acceptable classification accuracy for the DPI (Ghasemzadeh et al, 2025). Summary statistics were averaged for each group of three days. This process was repeated 1000 times to account for different random combinations of days within each group. The mean and 95% confidence intervals (CIs) were used for analysis. To address long-term reliability, 60 of the 125 subjects also had a second week of voice monitoring taken six months later. Summary statistics from the daily averages across all seven days were calculated for the same features and compared across both weeks. Both analyses were run on all voiced frames, as well as speech and singing voice only. A two-way, random effects Intraclass Correlation Coefficient was calculated to analyze agreement for the DPI, SPL skewness, and the standard deviation (SD) of the differences between the first two harmonics (H1-H2) for all voice, speech, and singing.

Results:

The within-week ICC results are listed in Table 1. The ICC for the DPI was .73 (95% CI: .67–.79) for all voicing. Reliability for speech was significantly higher than singing, but not for all voicing. Reliability was higher with H1-H2 standard deviation when compared to SPL skewness for all phonation and speech.

Regarding long-term reliability over six months, the ICC for all phonation was .70 (95% CI: .54–.81). The ICC for speech only was .76 (95% CI: .63–.85) and .64 (95% CI: .47–.77) for singing.

Conclusions:

Despite the expected variability in daily life, test-retest reliability for the DPI was in the moderate-to-good range for all phonation and speech when calculated within the same week and across six-months. Higher DPI reliability values on short- and long-term time scales may be possible

Table 1: Intraclass correlation coefficient (ICC) values documenting the test-retest reliability of the Daily Phonotrauma Index (DPI) and its component measures using within-week sampling of groups of three days.

		DPI	SPL skew	H1-H2 SD (dB)
All Voice	ICC	.73*	.63*	.81*
	95% CI	.67–.79	.55–.71	.76–.85
Speech	ICC	.75*	.67*	.82*
	95% CI	.70–.80	.60–.74	.77–.86
Singing	ICC	.65*	.62*	.67*
	95% CI	.57–.69	.54–.71	.60–.73

* p < .001

if monitoring is constrained to similar contexts (e.g., vocal demand) in daily life. DPI values from speaking voice data were more reliable than singing voice, potentially because the DPI was trained on mostly speaking voice data. The consistency in reliability between all voice and speaking is likely because the majority of phonation was speech in this sample. H1-H2 standard deviation is thought to capture vocal fold closure dynamics, whereas SPL skewness reflects the magnitude of phonatory forces towards higher values in daily life (Van Stan et al, 2020b). Since these participants had typical vocal folds throughout the monitoring time periods (and thus did not have to compensate for the presence of lesions), this may explain why H1-H2 standard deviation was more reliable than SPL skewness. Future work can focus on establishing thresholds of change for a phonotrauma early warning system in heavy voice users and a response-to-treatment threshold in patients with PVH.

References:

- Ghasemzadeh, H., Hillman, R. E., Van Stan, J. H., & Mehta, D. D. (2025). Effects of recording condition and number of monitored days on discriminative power of the daily phonotrauma index. *Journal of Speech, Language, and Hearing Research*, ePub Ahead of Issue, https://doi.org/10.1044/2024_JSLHR-24-00237
- 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. https://doi.org/10.1044/2020_AJSLP-20-00104
- Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of chiropractic medicine*, 15(2), 155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>
- Mehta, D. D., Van Stan, J. H., Zañartu, M., Ghassemi, M., Guttag, J. V., Espinoza, V. M., ... & Hillman, R. E. (2015). Using ambulatory voice monitoring to investigate common voice disorders: Research update. *Frontiers in bioengineering and biotechnology*, 3, 155.
- Nudelman, C. J., Ortiz, A. J., Fox, A. B., Mehta, D. D., Hillman, R. E., & Van Stan, J. H. (2022). Daily Phonotrauma Index: An objective indicator of large differences in self-reported vocal status in the daily life of females with phonotraumatic vocal hyperfunction. *American Journal of Speech-Language Pathology*, 31(3), 1412-1423.
- Trevethan, R. (2017). Intraclass correlation coefficients: clearing the air, extending some cautions, and making some requests. *Health Services and Outcomes Research Methodology*, 17(2), 127-143.
- Van Stan, J. H., Mehta, D. D., Ortiz, A. J., Burns, J. A., Toles, L. E., Marks, K. L., Vangel, M., Hron, T., Zeitels, S., & Hillman, R. E. (2020a). Differences in weeklong ambulatory vocal behavior between female patients with phonotraumatic lesions and matched controls. *Journal of Speech, Language, and Hearing Research*, 63(2), 372–384. https://doi.org/10.1044/2019_JSLHR-19-00065
- Van Stan, J. H., Mehta, D. D., Ortiz, A. J., Burns, J. A., Marks, K. L., Toles, L. E., ... & Hillman, R. E. (2020b). 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. *Journal of Speech, Language, and Hearing Research*, 63(12), 3934-3944.
- Van Stan, J. H., Ortiz, A. J., Marks, K. L., Toles, L. E., Mehta, D. D., Burns, J. A., Hron, T., Stadelman-Cohen, T., Krusemark, C., Muise, J., Fox, A. B., Nudelman, C., Zeitels, S., & Hillman, R. E. (2021). Changes in the daily phonotrauma index following the use of voice therapy as the sole treatment for phonotraumatic vocal hyperfunction in females. *Journal of Speech, Language, and Hearing Research*, 64(9), 3446–3455. https://doi.org/10.1044/2021_JSLHR-21-00082

Towards Reproducible Acoustic Measurements Obtained during Rigid Laryngoscopy with an Endoscope-Mounted Microphone

Jan G. Švec^{*1,2}, Dominika Valášková¹, Jitka Vydrová²

¹ Voice Research Lab, Dept. Experimental Physics, Faculty of Sciences, Palacký University, Olomouc, Czechia ² Voice and Hearing Centre Prague, Medical Healthcom, Ltd., Prague, Czechia

*Correspondence address: jan.svec@upol.cz

Keywords:

rigid laryngoscopy; acoustic voice recording; mouth-to-microphone distance

Abstract

Introduction & rationale: Glottal adjustments and vocal fold oscillations change with vocal intensity, pitch and quality (1-5). Therefore, it is desirable to capture and record the vocal sound during laryngoscopy. Nowadays, modern rigid endoscopes are equipped with embedded or endoscope-mounted microphones, making the acoustic voice recordings to be comfortably obtained simultaneously with laryngeal videos. The captured acoustic signals allow measuring the acoustic properties of voice. Some of measured acoustic properties, particularly the sound pressure level (SPL), are of limited use without knowing the mouth-to-microphone (MTM) distance, however (6). The MTM distance for such microphones depends on the placement of the microphone with respect to the tip of the endoscope, as well as on the depth of endoscope insertion into the mouth, and it varies among examined subjects and different rigid endoscopes. It is cumbersome to measure the MTM for each subject during laryngoscopic examinations. Therefore, we seek an easy and comfortable solution for determining the MTM distance in these cases so that the acoustic recordings and SPL measurements of voice obtained during laryngoscopy are reasonably accurate and reproducible.

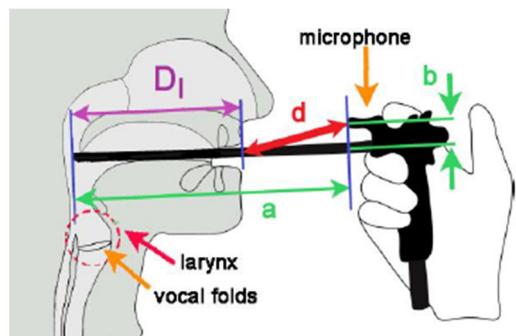


Figure 1. Insertion of a rigid laryngeal endoscope (laryngoscope) into the mouth to observe the larynx and the vocal folds during voice production in clinical voice examination. A small microphone attached to the laryngoscope captures the produced voice. The mouth-to-microphone (MTM) distance d , influences the properties of the captured sound. This distance depends on the position of the microphone with respect to the tip of the laryngoscope (parameters a and b) and the depth of insertion of the laryngoscope into the mouth D_l . Reprinted from Valášková et al., 2023 (7).

Objectives: We propose an empirical rule for determining the representative MTM distance (d) in adults for any 70° or 90° rigid laryngoscope with a fixed microphone using the known values of the depth of insertion of the laryngoscope into the mouth as a reference distance (D_l in Figure 1). We determine the uncertainty of the voice SPL at the representative MTM distance d (SPL@ d) using the known inter-individual variation of D_l .

Methods:

We use this empirical rule for determining the representative MTM distance d in adults:

$$d = \sqrt{(a-D_l)^2 + b^2}, \text{ (eq. 1)}$$

Here, a and b represent the horizontal and vertical position of the microphone with respect to the tip of the laryngoscope (see Figure 1) and the D_l values are 9.3 and 9.0 cm for the 70° and 90° laryngoscopes, respectively. This rule and the representative D_l values were recently determined by Valášková et al. (7) by measuring the insertion depth of rigid laryngeal endoscopes with 70° and 90° view in 60 adult women and 60 adult men.

Results:

For microphones fixed at the distances between 15 and 40 cm from the tip of the laryngoscope (distance a in Fig.1) and less than 10 cm above the laryngoscope (distance b in Fig.1), eq.1 reveals that the final representative MTM

distances d are between 5 and 35 cm from the lips, respectively. Considering the inter-individual differences in the laryngoscope insertion depth across adults, which were determined from the standard deviations of D_1 (± 0.9 cm and ± 0.7 cm for the 70° and 90° laryngoscopes, respectively), the standard uncertainties of these MTM distances are between 16 % and 2.5 %, respectively (7).

The inter-individual variability of the MTM distance d influences the accuracy of the measurement of vocal SPL@ d . For the 90° laryngoscope, the SPL uncertainty (95% confidence interval) gets smaller than ± 2 or ± 1 dB at the representative MTM distances longer than 7 or 14 cm, respectively. For the 70° laryngoscope, the SPL uncertainty gets smaller than ± 2 or ± 1 dB at the representative MTM distances longer than 9 or 17 cm, respectively.

Conclusions:

Since the SPL uncertainty below ± 2 dB is usually tolerable in acoustic measurements, representative MTM distances greater than 7-9 cm are suitable for measuring vocal sound with microphones fixed to rigid laryngeal endoscopes. Such representative MTM distances are achievable when the microphone is fixed at a position farther than 18 cm and 16 cm horizontally from the tip of the 70° and 90° laryngoscope, respectively. Our solution allows laryngologists and laryngoscope manufacturers to set and determine the representative MTM distance with reasonable accuracy for any rigid laryngeal endoscope with a microphone attached, avoiding the need for measuring this distance *in vivo* in routine practice. Reporting this distance allows the acoustic measurements to be more meaningful and reproducible, so that they are comparable across different institutes around the world. Relating the acoustic outputs to the laryngeal stroboscopic and high-speed videoendoscopic findings allows obtaining more insights into glottal function during voice production.

Acknowledgment:

The study was supported by the internal grant IGA_PrF_2024_030 from the Palacký University in Olomouc.

References:

1. Sulter AM, Schutte HK, Miller DG. Standardized laryngeal videostroboscopic rating: differences between untrained and trained male and female subjects, and effects of varying sound intensity, fundamental frequency, and age. *J Voice*. 1996;10(2):175-89. doi:10.1016/s0892-1997(96)80045-2
2. Woo P. Quantification of videostroboscopic findings--measurement of the normal glottal cycle. *Laryngoscope*. 1996;106 (suppl.no.79)(3, part 2):1-27. doi:10.1097/00005537-199603001-00001
3. Švec JG, Schutte HK, Šram F. Variability of vibration of normal vocal folds as seen in videokymography. In: Dejonckere PH, Peters HFM, editors. *Communication and its disorders: a science in progress Proceedings 24th Congress International Association of Logopedics and Phoniatrics*, Amsterdam, the Netherlands, August 23-27, 1998 Vol I: International Association of Logopedics and Phoniatrics; 1999. p. 122-5.
4. Yokonishi H, Imagawa H, Sakakibara K-I, Yamauchi A, Nito T, Yamasoba T, et al. Relationship of various open quotients with acoustic property, phonation types, fundamental frequency, and intensity. *J Voice*. 2016;30(2):145-57. doi:10.1016/j.jvoice.2015.01.009
5. Echternach M, Döllinger M, Köberlein M, Kuranova L, Gellrich D, Kainz MA. Vocal fold oscillation pattern changes related to loudness in patients with vocal fold mass lesions. *Journal of Otolaryngology - Head and Neck Surgery*. 2020;49(1). doi:10.1186/s40463-020-00481-y
6. Švec JG, Granqvist S. Tutorial and guidelines on measurement of sound pressure level in voice and speech. *Journal of Speech Language and Hearing Research*. 2018;61(3):441-61. doi:10.1044/2017_JSLHR-S-17-0095
7. Valášková D, Vydrová J, Švec JG. Determining the Mouth-to-Microphone Distance in Rigid Laryngoscopy: A Simple Solution Based on the Newly Measured Values of the Depth of Endoscope Insertion into the Mouth. *J Clin Med*. 2023;12(24):7560. doi:10.3390/jcm12247560

PODIUM SESSION 2

Tailored stimuli to investigate the neural coding of speech

Etienne Gaudrain¹, Sarah Verhulst², Deniz Başkent³

¹ Lyon Neuroscience Research Centre, Université Lyon 1, CNRS UMR5292, Inserm U1028, Lyon, France

² Department of Information Technology, Ghent University, Belgium

³ Dept. of Otorhinolaryngology, University Medical Center Groningen / University of Groningen, Netherlands

Keywords: (3 maximum)

Synthesis, Perception, Neuroscience

Introduction & rationale:

Speech perception relies on diverse neural coding mechanisms to cope with the richness and complexity of the speech signal. Understanding these mechanisms is particularly crucial for the rehabilitation of hearing loss. Typical psychoacoustical and neurophysiological approaches involve simple artificial stimuli with well controlled physical properties that can be modified at will by the experimenter. Such stimuli, like pure tones, have been used to characterise many auditory functions such as pitch and loudness perception (e.g. Moore, 2003). Pure tones are also used for audiometry, i.e., the clinical assessment of sensorineural hearing loss. However, the simplicity of these stimuli hinders the generalisation of the uncovered neural mechanisms to more complex and real-life related stimuli, such as speech.

Coincidentally, pure-tone audiometry is also unable to detect cochlear synaptopathy, a form of hearing impairment that is also referred to as “hidden hearing loss” (Liberman & Kujawa, 2017; Plack et al., 2014). Cochlear synaptopathy results from a loss of functional connectivity between the hair cells in the cochlea and the auditory nerve. While this impairment does not affect the detectability of pure tones, it is thought to affect speech coding, and in particular, the ability to cope with noisy environments. Previous research has identified that neural mechanisms that specifically encode the *temporal fine structure* (TFS) of the sound played a central role in the ability to understand speech in noisy environments (Lorenzi et al., 2006), and these mechanisms have been found to be directly affected by cochlear synaptopathy in animal studies (Kujawa & Liberman, 2009).

To better understand the effects of cochlear synaptopathy on the neural coding of speech — and to better understand the neural coding of speech in general — large consortium studies gathering empirical evidence from both humans and animals are necessary. To target TFS coding, speech stimuli can be designed to specifically differ in how much TFS coding they promote. These stimuli can then be used to collect behavioural evidence in humans, and electrophysiological evidence in animals.

Objectives:

The purpose of the presented study is to produce realistic speech stimuli that are as tightly controlled in their physical properties as pure tones, while remaining ecologically valid. The acoustic features supporting the involved phonetic contrasts must either target TFS coding, or not. Furthermore, because the human studies are to be conducted as part of an international consortium to produce a large data set, the stimuli are to be compatible with German, Dutch and French. Finally, because TFS may be more crucial for the coding transients than for steady-state spectral information, both vowels and consonants are considered.

Methods:

To target TFS coding, we opted for a discrimination task on pairs of phonemes that differ by a single acoustic trait that is either potentially coded through TFS, or not. TFS coding is limited by neural phase-locking, which only takes place up to a given frequency. Although there is much debate about what the exact limit of phase-locking is in humans, some consensus seems to converge towards a value of 1.5 kHz (Verschooten et al., 2019), which we chose for the present purpose. We therefore looked for pairs of phonemes that are differentiated by a single acoustic trait that is either below, or above this 1.5 kHz limit. When the trait is above 1.5 kHz, it cannot be coded with TFS. However, when the trait is below 1.5 kHz, it can potentially be coded using TFS.

The first phase consisted of selecting vowel and syllable pairs from the phonetic inventories of German, Dutch and French that exist in all three languages. From the overlapping items, we selected pairs differing primarily in their first formant, and pairs differing primarily in their second formant. This resulted in the items shown in Table 1, along with examples of words from the three languages where these phonemes are used.

The selected items were then recorded as syllables, but without a word carrier, and then processed using analysis-resynthesis to isolate the discrimination features to be strictly below or above the phase-locking limit, and to create morphed continua between the two items of each pair. Using the WORLD vocoder (Morise et al., 2016), the recorded tokens were decomposed into an f_0 contour, an aperiodicity matrix, and a spectral envelope. For the vowels, the spectral envelope was approximated using Gaussian mixtures. The morphing was then generated by translating the Gaussian components located on one side or the other of the 1.5 kHz limit. For the CV-syllables, a spectrotemporal window was defined (<100 ms, < or >1.5 kHz), where differences between items within a pair were allowed. Outside of this window, the spectral envelope was made exactly the same. Before 100 ms, the neutralised part consisted of the average spectral envelope of the two items of the pair. After 100 ms, in the steady-state part of the syllable, the exact same vowel was used as the isolated vowel presented previously.

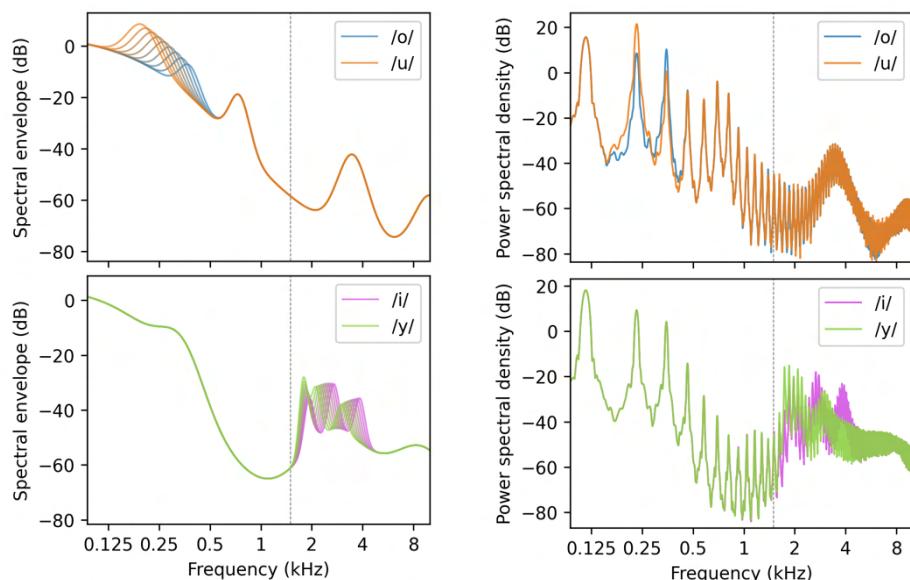
Table 1 — Selected vowel and syllable contrasts and example words in German, French and Dutch.

		German	French	Dutch
Vowels	<1.5 kHz	/o/ oder [or]	ôter [to remove]	oké [okay]
	/u/	Universität [university]	outil [tool]	oester [oyster]
Syllables	>1.5 kHz	/i/ ikonisch [iconic]	ignorant [ignorant]	iets [something]
	/y/	über [over]	université [university]	universiteit [university]
<1.5 kHz	/du/	Du [you]	doux [soft]	doe [do]
	/bu/	Butan [butane]	bout [end]	boete [fine]
>1.5 kHz	/di/	die [the]	dit [said]	die [that]
	/bi/	Bier [beer]	bipède [biped]	bier [beer]

Results:

Spectral envelopes and average spectra of the generated stimuli are shown in Figure 1.

Figure 1 — Spectral envelopes of the 9-step continua between vowels (left panels) and average spectra of the resynthesised realisations of the extreme steps in the continuum (right panels). The vertical dashed line shows the 1.5 kHz phase-locking limit.



Conclusions:

Behavioural discrimination data and physiological recordings in animals and humans were obtained for these stimuli, confirming that the targeted mechanisms are indeed involved. Overall, the approach demonstrates that analysis-resynthesis can be used to characterise specific neural speech coding mechanisms with a precision comparable to that obtained with simplified stimuli, while retaining ecological validity for communication.

References:

- Kujawa, S. G., & Liberman, M. C. (2009). Adding insult to injury: Cochlear nerve degeneration after “temporary” noise-induced hearing loss. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 29(45), 14077–14085. <https://doi.org/10.1523/JNEUROSCI.2845-09.2009>
- Liberman, M. C., & Kujawa, S. G. (2017). Cochlear synaptopathy in acquired sensorineural hearing loss: Manifestations and mechanisms. *Hearing Research*, 349, 138–147. <https://doi.org/10.1016/j.heares.2017.01.003>
- Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., & Moore, B. C. J. (2006). Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. *Proceedings of the National Academy of Sciences of the United States of America*, 103(49), 18866–18869. <https://doi.org/10.1073/pnas.0607364103>
- Moore, B. C. J. (2003). *An introduction to the psychology of hearing* (Fifth edition). Academic Press.
- Morise, M., Yokomori, F., & Ozawa, K. (2016). WORLD: A vocoder-based high-quality speech synthesis system for real-time applications. *IEICE Transactions on Information and Systems*, E99.D(7), 1877–1884. <https://doi.org/10.1587/transinf.2015EDP7457>
- Plack, C. J., Barker, D., & Prendergast, G. (2014). Perceptual consequences of “hidden” hearing loss. *Trends in Hearing*, 18, 2331216514550621. <https://doi.org/10.1177/2331216514550621>
- Verschooten, E., Shamma, S., Oxenham, A. J., Moore, B. C. J., Joris, P. X., Heinz, M. G., & Plack, C. J. (2019). The upper frequency limit for the use of phase locking to code temporal fine structure in humans: A compilation of viewpoints. *Hearing Research*, 377, 109–121. <https://doi.org/10.1016/j.heares.2019.03.011>

Sensorimotor control of pitch across age in adulthood

Valentine Lucquiault¹, Thomas Tienkamp^{1,2,3}, Hedwig Sekeres¹, Martijn Wieling¹,
Defne Abur^{1,3}

¹ Center for Language and Cognition Groningen, University of Groningen, Groningen, the Netherlands

² Department of Oral and Maxillofacial Surgery, University Medical Center Groningen, Groningen, the Netherlands

³ Research School of Behavioral and Cognitive Neurosciences, University of Groningen, Groningen, the Netherlands

Keywords: vocal motor control, pitch acuity, pitch reflex

Abstract

Introduction & rationale:

Sensorimotor control of speech relies on perception (auditory and somatosensory feedback) and production working together to achieve fluent speech output. Auditory-motor control of speech involves primarily using auditory feedback to monitor and produce the intended speech output as posited by the Directions Into Velocities of Articulators (DIVA) model [1]. In this framework, auditory feedback control mechanisms are responsible for detecting errors in speech and for triggering subsequent motor corrections. Empirically, experimental findings show that sudden, unexpected changes in auditory feedback during speech prompt a reflexive correction during speech production. For example, when auditory feedback of vocal pitch is suddenly shifted upwards, speakers respond by lowering their pitch ("pitch reflex") and opposing the direction of the change [2]. "Over-corrections" of the pitch reflex have been observed in older adults compared to younger adults, which suggest reduced vocal control in the aging process [3]. Previous research suggests that the robustness of these reflexive responses may be linked to perceptual acuity to same percept [4]. Further, studies in typical speakers found that perceptual acuity to speech parameters such as pitch can decrease with age [5] and are affected by musical experience [6].

However, it remains unclear whether the relationship between vocal control (as measured by the pitch reflex) and perceptual acuity to self-generated voice changes with age in adulthood. Clarifying how sensorimotor control is impacted by age, both in terms of sensory feedback control and voice production, is crucial to have a baseline for comparison when assessing speech disorders than occur in older age and that have sensorimotor disruptions (e.g., Parkinson's disease, [7]).

Objectives:

The aims of the current work were to assess the following metrics across age in adulthood: (1) perceptual acuity for self-generated pitch; (2) vocal pitch shift reflex responses; (3) the relationship between acuity and pitch shift reflex responses.

Methods:

A total of 120 native speakers of Italian (59 males, 61 females) aged 16 to 74 years (age mean $M = 40.0$, age standard deviation $SD = 15.3$) participated in the study. Participants provided informed consent for the study in accordance with institutional ethical approval. Exclusion criteria were having a native language other than Italian, a history of speech, language, hearing, or neurological disorder, and not meeting the hearing screening requirements. Demographic information of the participants is presented on Figure 1.

Prior to the experiment, participants answered a questionnaire about their demographic and language profile: age, sex, native language, history of speech and hearing disorders, musical and singing experience. Answers to the questionnaire were used to assess eligibility and singing status, with participants having five or more years of singing experience being defined as singers. Next, participants completed a hearing screening to ensure typical hearing for their age (using pulsed pure tones delivered via MAICO MA25 audiometer). Age-appropriate hearing thresholds were set based on [8]: individuals under the age of 50 were required to hear all frequencies at 25 dB HL; participants over the age of 50 years were included if they heard the tones at 25 dB HL for 250 Hz and 500 Hz, and at 40 dB HL for 1000 Hz and above. All experimental data were collected in a sound-attenuated booth in a mobile laboratory [9]. Visual cues for sustaining phonation were displayed on a monitor located one meter away from the participant chair. Participants were fitted with an omni-directional microphone (Shure MX153) at 7 cm from the mouth and over-the-ear headphones (Sennheiser 280 Pro HD). Speech was recorded at a sampling frequency of 44.1 kHz and digitized via a Microbook IIc (MOTU) soundcard.

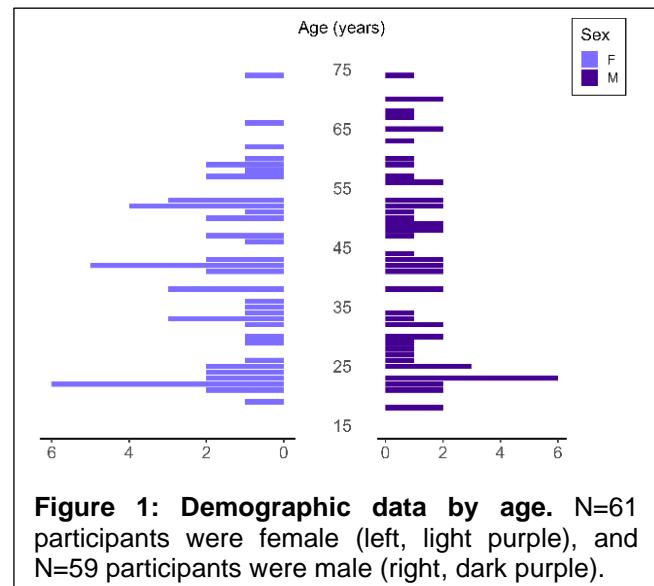


Figure 1: Demographic data by age. N=61 participants were female (left, light purple), and N=59 participants were male (right, dark purple).

Participants first performed a two-way forced-alternative perceptual pitch discrimination task, to determine their pitch acuity threshold, then they completed the vocal pitch reflex task. For the acuity task, participants were asked to sustain /a:/ for three seconds. The most stable 500-ms medial interval of their production was then extracted and used as reference stimulus for the discrimination task. In each trial, participants heard two stimuli (in a randomised order): one was the reference stimulus consisting of their original voice recording, the other was either the same stimulus, or a pitch-shifted version of their voice in terms of f_0 (20% of trials). The magnitude of the pitch shift was modified after each trial based on the participant's answer: two correct consecutive answers caused the f_0 difference in the following trial to 0.04 semitones, whereas one incorrect answer caused the f_0 difference to be 0.04 semitones larger. The task lasted for ten direction reversals (i.e. ten increases or decreases in the voice f_0 difference). For the pitch reflex task, participants were asked to sustain the vowel /a:/ for three seconds (cued by on-screen visual displays) for 56 trials while receiving real-time feedback of their production through the headphones. For 20% of the 56 trials selected randomly, the auditory feedback received through the headphones was shifted +1 semitone above the participant's vocal f_0 from the microphone signal. All pitch shifts were applied using the Audapter software [10] and custom MATLAB [11] scripts.

Results and discussions:

All data has been collected and analyses are ongoing. The full analyses will be completed and presented at the conference. To date, we have completed the data analysis of the acuity task (Figure 2). Pitch acuity thresholds were computed by averaging the f_0 difference (in semitones) across the last six reversals (i.e., the moments in the task when the difference in f_0 changed). Based on the seemingly non-linear pattern in Figure 2, subsequent statistical analysis will rely on Generalised Additive Models (GAMs; [12, 13]) that we will build using R Statistical Software version 4.4.0 [14] in RStudio [15]. Data analysis for the pitch reflex will focus on response magnitude (in semitones) and will be conducted using an automated, custom MATLAB script. We will then compare the magnitude of the responses to the vocal pitch shift reflex (in semitones) to the perceptual acuity for pitch (in semitones). The effect of age, sex, and singing experience will also be considered in the analyses given previous work showing influences of these variables on sensorimotor voice control.

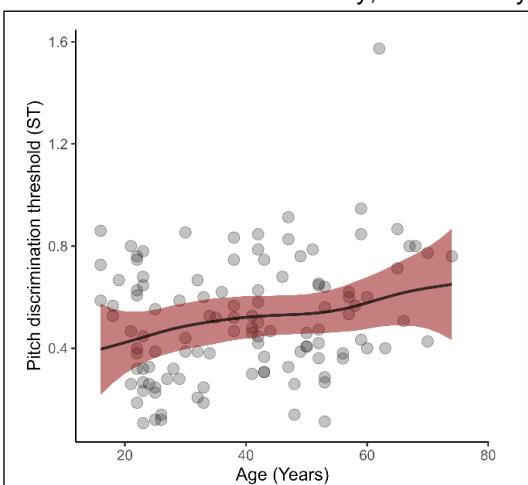


Figure 2: Pitch discrimination threshold (ST=semitones) plotted against age (in years). Shaded region represents the confidence interval for the effect of age on discrimination threshold.

References:

- [1] Tourville, J. A., & Guenther, F. H. (2011). The DIVA model: A neural theory of speech acquisition and production. *Language & Cognitive Processes*, 26(7), 952–981.
- [2] Burnett, T. A., Freedland, M. B., Larson, C. R., & Hain, T. C. (1998). Voice F0 responses to manipulations in pitch feedback. *The Journal of the Acoustical Society of America*, 103(6), 3153-3161.
- [3] Liu, H., Russo, N. M., & Larson, C. R. (2010). Age-related differences in vocal responses to pitch feedback perturbations: A preliminary study. *The Journal of the Acoustical Society of America*, 127(2), 1042-1046.
- [4] Smith, D. J., Stepp, C., Guenther, F. H., & Kearney, E. (2020). Contributions of auditory and somatosensory feedback to vocal motor control. *Journal of Speech, Language, and Hearing Research*, 63(7), 2039-2053.
- [5] Fulton, K. S. (2007). *Vocal efficiency in trained singers vs. non-singers*. Brigham Young University.
- [6] Broadfoot, C. K., Abur, D., Hoffmeister, J. D., Stepp, C. E., & Ciucci, M. R. (2019). based updates in swallowing and communication dysfunction in Parkinson disease: implications for evaluation and management. *Perspectives of the ASHA special interest groups*, 4(5), 825-841.
- [7] Schow, R. L. (1991). Considerations in selecting and validating an adult/elderly hearing screening protocol. *Ear and hearing*, 12(5), 337-348.
- [8] Wieling, M., Rebernik, T., & Jacobi, J. (2023). SPRAAKLAB: a mobile laboratory for collecting speech production data. In *Proceedings of the 20th International Congress of Phonetic Sciences* (pp. 2060-2064). Guarant International.
- [9] Cai, S., Boucek, M., Ghosh, S. S., Guenther, F. H., & Perkell, J. S. (2008). A system for online dynamic perturbation of formant trajectories and results from perturbations of the Mandarin triphthong /iau/. In *Proceedings of the 8th International Seminar on Speech Production in Strasbourg, France*, (pp. 65–68).
- [10] MATLAB. (2024). version 23.2 (R2023b). Natick, Massachusetts: The MathWorks Inc. Available: <https://www.mathworks.com>
- [11] Wieling, M. (2018). Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics*, 70, 86-116.
- [12] Wood, S. N. (2017). *Generalized additive models: an introduction with R*. Chapman and Hall/CRC.
- [13] R Core Team (2024). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available: <https://www.R-project.org/>.
- [14] RStudio Team (2024). Rstudio: Integrated development environment for r. RStudio, PBC, Boston, MA, USA. Available: <http://www.rstudio.com/>

Voice loudness, self-assessment and auditory-sensory feedback in voice production in Parkinson's disease

Francisco Contreras-Ruston^{1,2,3}, Antoni Callen⁴, Matías Zañartu⁵, Héctor Arriagada-Concha⁶, Antonia Lagos-Villaseca⁷, Sandra Rojas⁸, Carla Napolitano⁷, Jordi Navarra¹, Sonja Kotz²

¹Department of Cognition, Development and Educational Psychology, University of Barcelona, Barcelona, Spain

²Faculty of Psychology and Neuroscience, Department of Neuropsychology & Psychopharmacology, Maastricht University, Maastricht, The Netherlands.

³Speech-Language Pathology and Audiology Department – Universidad de Valparaíso, San Felipe, Chile

⁴Institut de Recerca, Hospital Sant Joan de Déu, Barcelona, Spain

⁵Universidad Técnica Federico Santa María, Viña del Mar, Chile

⁶Faculty of Health, University of Santo Tomás and speech therapist at Clínica Dávila, kinesiotherapy department

⁷Department of Otolaryngology, Faculty of Medicine, Pontificia Universidad Católica de Chile, Santiago, Chile

⁸Escuela de Fonoaudiología, Facultad de Odontología y Ciencias de la Rehabilitación, Universidad San Sebastián, Santiago, Chile

Keywords: *Parkinson's disease, voice, dysphonia, self-assessment-loudness, auditory-sensory feedback*

Abstract

Introduction / Objectives:

Parkinson's disease (PD) is frequently associated with symptoms, such as reduced vocal loudness. This vocal impairment in individuals with PD (IwPD) may stem from their response to auditory feedback errors, as suggested by Huang et al. (2019). Research on healthy individuals, such as that by Guenther et al. (2006), has highlighted the role of feedforward systems in voice regulation. However, the function of these mechanisms in IwPD remains less well understood. Additionally, the IwPD may not fully acknowledge the extent of their vocal difficulties. Self-awareness of one's voice is a critical aspect in clinical voice assessments, where clinicians often employ self-assessment tools to evaluate voice function in IwPD. Nevertheless, owing to reduced self-perception, IwPD may perceive their voice as functioning normally when using such tools. Electrophysiological studies have sought to address this challenge by testing auditory feedback, feedforward control, and motor suppression models during voice production. Knolle et al. (2019) utilized Motor-Induced Suppression (MIS) to examine event-related potentials (ERPs), such as N1 and P2, to investigate these mechanisms. The MIS phenomenon in IwPD has been linked to difficulties in controlling vocal loudness (Emmendorfer et al. 2021; Abur et al. 2018; Li et al. 2021).

Methods:

To explore the mechanisms underlying voice loudness feedback in IwPD, we conducted two studies.

Study 1

This study assessed voice self-perception and awareness by PROMs in three groups: 27 individuals with Parkinson's disease (IwPD, aged 61–79 years), 25 individuals with general voice disorders (GVD, aged 57–83 years), and 28 healthy controls (HC, aged 60–80 years). Univariate and multivariate analyses were used to compare the outcomes of these measures among the three groups.

Study 2

The study examined feedback mechanisms of voice loudness in IwPD using ERPs (N1 and P2) and the Button-Press Paradigm (BP) with three conditions: (a) Auditory Motor Condition (AMC) with participants tapping while speaking at +15dB above habitual loudness, (b) Auditory Only Condition (AOC) where participants listened to their voice at +15dB without tapping, and (c) Motor-Only Condition (MOC) where participants tapped without voice or sound. Each condition included 100 trials. A multimodal design with three cross-sectional studies will be used to compare healthy individuals with IwPD.

Results:

Study 1

The IwPD group exhibited significantly reduced self-perception and awareness of voice problems compared with the GVD and HC groups, despite showing slightly higher PROMs scores than the HCs. Notably, in the IwPD group, age did not have a significant impact on self-perception or awareness of voice issues. In contrast, the GVD group demonstrated pronounced differences across all PROMs and voice parameters when compared with both the HC and IwPD groups. Principal Component Analysis (PCA) further highlighted significant intergroup differences in total scores and voice loudness, underscoring distinct patterns of voice self-awareness and perception among the groups.

Study 2

The efference copy in IwPD is impaired, leading to compromised suppression mechanisms, which is reflected in the ERPs. This alteration is evident both at habitual loudness and +15 dB, where ERP amplitudes differ significantly from those observed in healthy individuals.

Conclusions:

This study highlights key differences in voice perception and production in IwPD. In Study 1, it shows that IwPD reports fewer vocal symptoms than other voice disorders, but also less than healthy controls, indicating a specific discrepancy in self-perception of voice problems and PD. Study 2 revealed alterations in the auditory-sensory feedback mechanism in IwPD, evidenced by changes in ERP amplitudes, suggesting impaired neural processing for voice loudness control. These findings emphasize the need for tailored assessment and treatment approaches in voice therapy for PD taking into account both their subjective perceptions and objective neurophysiological changes.

References:

- Guenther, F. H. (2006). Cortical interactions underlying the production of speech sounds. *J. Commun. Disord.* 39, 350–365. doi: 10.1016/j.jcomdis.2006.06.013
- Knolle, F., Schwartze, M., Schröger, E., & Kotz, S. A. (2019). Auditory predictions and prediction errors in response to self-initiated vowels. *Frontiers in Neuroscience*, 13, 1146. <https://doi.org/10.3389/fnins.2019.01146>
- Emmendorfer, A. K., Bonte, M., Burnyte, B., Kotz, S. A (2020). Phonological and temporal regularities lead to differential ERP effects in self-and externally generated speech. *Scientific Reports*, 10(1), 9917. <https://doi.org/10.1101/2021.05.04.442414>
- Abur, D., Lester-Smith, R.A., Daliri, A., Lupiani, A.A., Guenther, F.H., Stepp, C.E., 2018. Sensorimotor adaptation of voice fundamental frequency in Parkinson's disease. *PLoS One* 13. <https://doi.org/10.1371/journal.pone.0191839>
- Li, Y., Tan, M., Fan, H., Wang, E.Q., Chen, L., Li, J., Chen, X., Liu, H., 2021. Neurobehavioral Effects of LSVT® LOUD on Auditory-Vocal Integration in Parkinson's Disease: A Preliminary Study. *Front Neurosci* 15. <https://doi.org/10.3389/fnins.2021.624801>

PODIUM SESSION 3

Deep Learning-Based Laryngo-Tracheal Landmark Tracking for Rapid and Safe Vocal Fold Injection

Amulya Vankayalapati¹, Nour Awad¹, Manikanta Varaganti², Laura J. Brattain², Gregory R. Dion¹

¹ Department of Otolaryngology-Head and Neck Surgery, University of Cincinnati College of Medicine, Cincinnati, OH, USA

² Department of Computer Science, University of Central Florida, Orlando, FL, USA

³ Department of Internal Medicine, University of Central Florida College of Medicine, Orlando, FL, USA

Keywords: Airway Management, Deep Learning, Vocal Fold, Neck Ultrasound

Abstract

Introduction & rationale: Ultrasound imaging has emerged as a valuable modality in airway management due to its speed, accessibility, and non-invasive nature. In addition to identifying and evaluating head and neck structures, ultrasound guidance plays an integral role in procedures such as cricothyrotomy, tracheotomy, and transcutaneous vocal fold injections (Lakhal, 2017; Patel, 2021). Ultrasound detection of key anatomical landmarks for these procedures is far superior to traditional palpation of the neck (Rudas, 2014). Despite its clinical benefits, ultrasound is limited by the considerable variability in patient body habitus and the technical proficiency of operators. The integration of an artificial intelligence (AI) model enhances this approach by providing real-time landmark detection for key anatomical structures, including the tracheal rings, cricoid cartilage, thyroid gland, vocal folds, and relevant vasculature. Our past projects utilized AI to detect anatomical structures using ultrasound in patients with both normal and complex anatomy (including patients with elevated BMI and laryngotracheal stenosis), supporting its efficacy in anatomical detection within a broad patient population (Khodagholi, 2024). Coupled with this real-time ultrasound guidance technology, we prototyped a semi-automated needle injection device to simplify and improve reliable, definitive airway management (**Figure 1**). The goal is to leverage our prior experience in real-time AI detection of neck anatomical structures and semi-autonomous device creation and miniature robotics to deliver a needle to the paraglottic space to assist with procedures such as vocal cord injections.

Objectives: This study leverages real time tracking and identification of vocal fold and laryngo-tracheal landmarks through a deep learning system to allow for semi-autonomous delivery of therapeutics to the vocal folds and paraglottic space.

Methods: Ultrasound (US) images of the neck were collected from 11 human subjects (4 male/7 females, avg age 52.6 ± 14.5) using a Terason uSmart 3200T system (Teratech, Burlington, MA) equipped with a 15L4A linear probe at a 4 cm imaging depth. Ultrasound transverse views of the anterior neck were collected as 10-second cineloops from subjects with the ultrasound probe moving cranially to caudally from the hyoid bone to the suprasternal notch, with particular focus at the level of the vocal folds. Placement accuracy was ensured with subject vocalization. Airway specialists annotated the scans using Yolo-Mark, creating bounding boxes and class labels of strap muscles, tracheal rings, thyroid gland, cricoid cartilage, thyroid cartilage, and vocal folds to facilitate model training (Bochkovskiy, 2020). The deep learning object detector YOLOv9 was then fine-tuned using Leave-One-Out cross-validation on the 11-subject neck ultrasound data. Several metrics including Area Under the Curve (AUC), F1 score, and mAP50 are computed to evaluate the accuracy of landmark detection in this dataset. Since there are very few vocal folds and tracheal ring images in each sweep when compared to the number of strap muscles and thyroid gland, we also employed class aware sampling and stable diffusion models to generate synthetic images thus mitigating the severe class imbalance.

Results: Table 1 shows the YOLOv9 performance metrics of using 3 diffusion models including baseline, repeat factor sampling, and class aware sampling models. Class Aware yielded the best results. **Figure 2** shows the original examples of tracheal ring (right) and vocal fold (left) ultrasound annotations (top) with corresponding Precision-Recall curves across different configurations (bottom).

Table 1. Performance metrics of baseline, repeat factor sampling, and class aware sampling models, including mAP50 and F1 scores.

Model	mAP@0.5	4-fold mAP@0.5	F1-score
Baseline	0.947	0.94	0.89
Repeat Factor	0.961	0.93	0.87
Class Aware	0.982	0.97	0.94

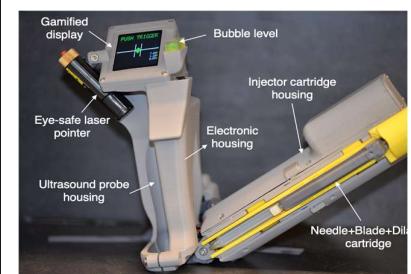


Figure 1. Real-time AI semi-autonomous Airway Access Systm.

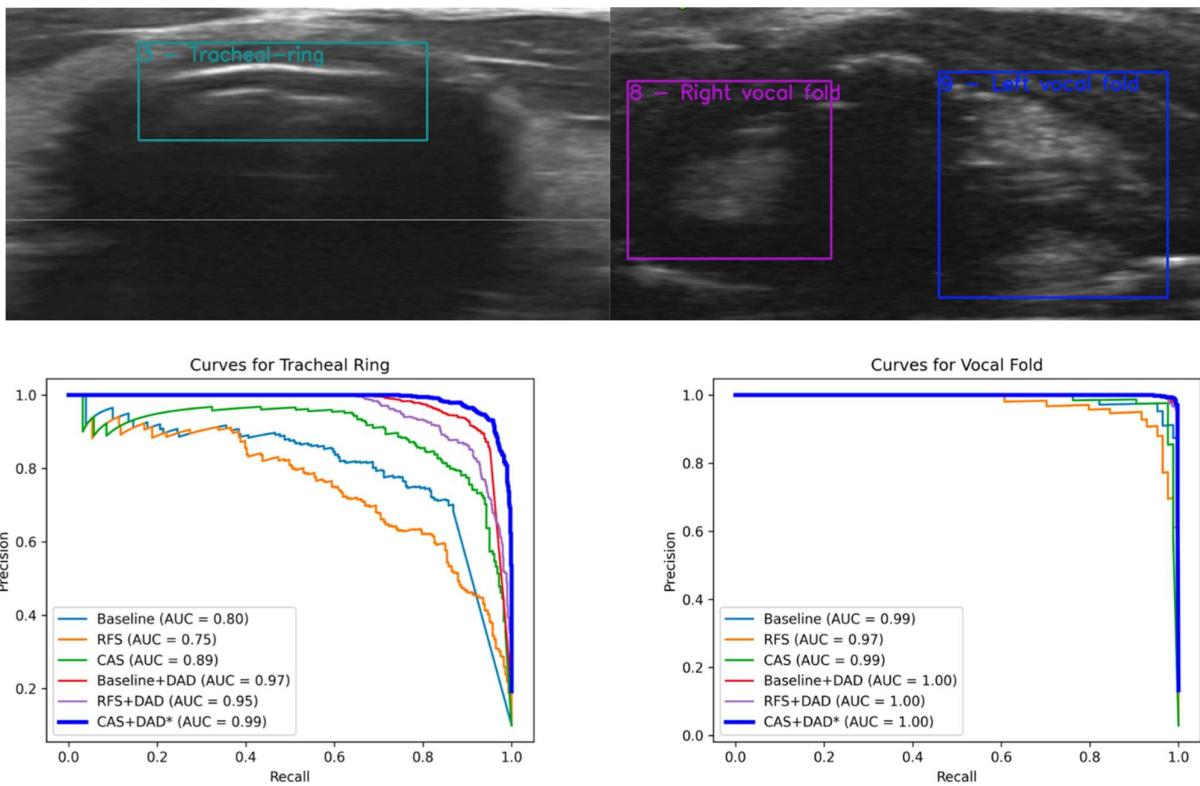


Figure 2. Original examples of tracheal ring (right) and vocal fold (left) ultrasound annotations (top) with corresponding Precision-Recall curves across different configurations (bottom).

Discussion and Conclusions: The integration of diffusion-based augmentation further enhances model performance, particularly for minority classes (e.g., tracheal ring and vocal fold). This suggests that strategic data augmentation combined with appropriate sampling mechanisms significantly improves detection performance for sparsely represented classes. Accurate detection and tracking are crucial to safe vocal fold injection. The use of AI in detecting anatomical landmarks on ultrasound can reduce risk and promote safe airway management through alerting clinicians to specific structures in the neck. Additionally, this tool may increase the consistency of procedures between providers, providing critical guidance in various settings. Based on our improved detection of anatomical landmarks in real time and inclusion of vocal folds and other laryngeal structures, these results are promising in the ability to adapt our semi-autonomous, AI-guided device for vocal fold and paraglottic space injection.

References:

- Bochkovskiy, A., Wang, C., & Liao, H. (2020). YOLOv4: Optimal Speed and Accuracy of Object Detection. *ArXiv*. <https://doi.org/10.48550/arXiv.2004.10934>.
- Khodagholi, Z., Sun, J., Awad, N., Vankayalapati, A., Dion, G., & Brattain, L. (2024). ULTRA-AIR: Ultrasound Landmark Tracking for Real-Time Anatomical Airway Identification and Reliability Check. *2024 IEEE 20th International Conference on Body Sensor Networks (BSN)*, Chicago, IL, USA, 2024, pp. 1-4, <https://doi.org/10.1109/BSN63547.2024.10780557>.
- Lakhal, K., Delplace, X., Cottier, J. P., Tranquart, F., Sauvagnac, X., Mercier, C., Fusciardi, J., & Laffon, M. (2007). The feasibility of ultrasound to assess subglottic diameter. *Anesthesia and analgesia*, 104(3), 611–614. <https://doi.org/10.1213/01.ane.0000260136.53694.fe>.
- Patel, S., Goyal, N., & Gniady, J. P. (2021). The role of ultrasound in various office-based laryngeal procedures. *Operative Techniques in Otolaryngology-Head and Neck Surgery*, 32(1). <https://doi.org/10.1016/j.otot.2020.11.001>
- Rudas, M., Seppelt, I., Herkes, R., Hislop, R., Rajbhandari, D., & Weisbrodt, L. (2014). Traditional landmark versus ultrasound guided tracheal puncture during percutaneous dilatational tracheostomy in adult intensive care patients: a randomised controlled trial. *Critical care (London, England)*, 18(5), 514. <https://doi.org/10.1186/s13054-014-0514-0>.

Mechanisms of vocal fold inferior medial bulging with thyroarytenoid muscle activation

Tsukasa Yoshinaga^{1,2}, Zhaoyan Zhang²

¹ Department of Mechanical Science & Biomechanics, Osaka University, Toyonaka, Osaka, Japan

² Department of Head and Neck Surgery, University of California, Los Angeles, California, USA

Keywords: numerical modeling, thyroarytenoid muscles, vocal fold medial bulging

Abstract

Introduction & rationale: The activation of the thyroarytenoid (TA) muscles has been shown to cause the inferior portion of the medial surface to bulge toward the glottal midline and increased the medial surface vertical thickness in canine larynges by Hirano [1]. This ability to modulate the vertical thickness has been shown to be essential to the control of voice quality [2]. More recent experiments with excised larynges also showed the inferior bulging with increasing activation of the TA (e.g., [3]). However, the mechanism of how TA muscle contraction causes inferior medial bulging of the vocal fold medial surface is still unclear.

In previous studies, various computational models of the human larynx have been proposed to clarify the functions of each intrinsic muscle, e.g., [4-5]. These studies showed that TA activation increased the fundamental frequency of the vocal fold vibration by stiffening the muscle fibers. However, these simulations did not demonstrate clear inferior bulging and the mechanisms remained unclear.

Objectives: The goal of this study is to investigate the mechanisms of vocal fold inferior medial bulging due to TA muscle activation in computational models of laryngeal muscle activation in two human larynges. By parametrically changing TA muscle fiber parameters, we examine the cause of vocal fold medial budging and changes in the vertical thickness of vocal folds with TA activation.

Methods: The geometry of three cartilages (arytenoid, cricoid, and thyroid cartilages), five intrinsic muscles (thyroarytenoid, lateral and posterior cricoarytenoid, cricothyroid, and interarytenoid muscles), and a cover layer of human larynx was obtained from an MRI scan of hemi-larynges fixed in a cylindrical container at the resting position. Larynges from a 57-year-old male and a 82-year-old female were scanned in a Bruker BioSpec 7 Tesla MRI (Bruker Biospin GmbH, Rheinstetten, Germany). The three-dimensional laryngeal structures were segmented and reconstructed from the MRI images, as shown in Fig. 1.

The muscle activation simulations were conducted by solving the equilibrium equations with a nonlinear finite deformation formulation. Each muscle was modeled as a nonlinear material with a passive and an active component. The tissues of TA and cover layer were modeled as an anisotropic Gasser–Ogden–Holzapfel (GOH) material, with stiffness different between the anterior-posterior and medial-lateral directions. The material constants were estimated by fitting the stress-strain curves measured using excised larynges [6]. All other laryngeal muscles were assumed to be an isotropic hyperelastic material and modeled as a nearly incompressible Yeoh model as described in [4]. Because the cartilages are much stiffer than the vocal folds and muscles and are expected to exhibit small deformations, they were modeled as a linear elastic material with a Young's modulus of 10 MPa and a Poisson's ratio of 0.47.

The active component of stress tensor for the TA muscle was modeled by including an active component in the constitutive model of the TA in the same way as [4], with the anisotropic stress applied along the TA muscle fiber direction. In this study, we parametrically varied the fiber angle of the TA muscle with respective to the horizontal glottal midline from 0° (along axis-x or glottal midline) to 45° on the x-z plane. The vertical thickness of vocal folds at the mid-membranous location at each activation level of TA muscle was calculated based on [7] to evaluate the degree of inferior medial bulging. The numerical simulations were conducted using a commercial software COMSOL ver. 6.1.

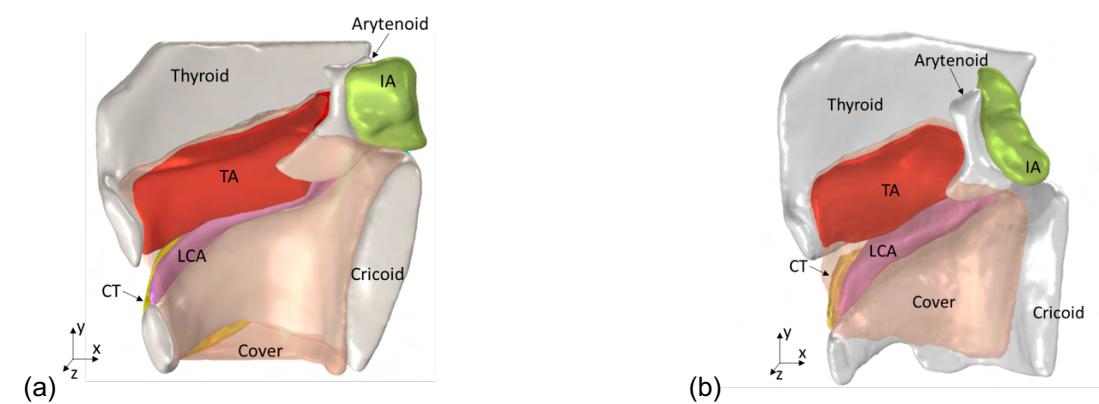


Fig. 1. Medial view of the MRI-based three-dimensional model of a male (a) and female (b) larynx.

Results: Changes in the vertical thickness of vocal folds at different fiber angles and TA muscle activation levels are plotted in Fig. 2. In general, the vertical thickness increased with increasing TA activation level. This increase is the largest for fiber angles between 17° to 31° in the male larynx, in which the vertical thickness increased from 2.5 to 5.2 mm when the TA activation level increased from 0 (no activation) to 1 (full activation). In comparison, the vertical thickness of the female larynx increased from 4.3 to 5.2 mm when the fiber angle was 31°. The medial-lateral displacement of the vocal folds within a mid-coronal plane is color coded and plotted in Fig. 3 for four fiber angles in the male larynx. The solid lines indicate the initial resting position. For the two intermediate fiber angles at 17° and 31°, the inferior part of the vocal fold medial surface moved towards the medial plane noticeably more than the superior part of the medial surface, increasing the vertical thickness of the vocal folds. This range of fiber angles is slightly larger than the angle formed by the medial edge of the TA muscle and the glottal midline, which was 9° and 17° at rest in the male and female larynx, respectively. Further analysis showed that at fiber angles above 17°, the TA muscle activation caused a rotation motion of the arytenoid around the vertical axis (y-axis), which resulted in the inferior medial bulging of the TA muscle.

Conclusions: Our simulations demonstrated that the orientation of the TA muscle fibers is important in determining the degree of inferior medial bulging and vertical thickness of the vocal fold medial surface, which is essential to voice quality control as shown in [2].

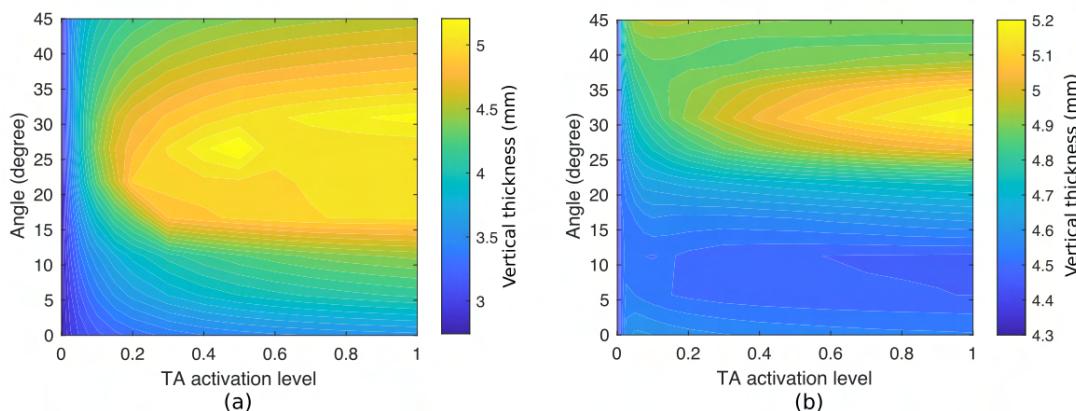


Fig. 2. Contour of vertical thickness at different TA activation levels (0: no activation; 1: full activation) and fiber angles for male (a) and female (b) larynges.

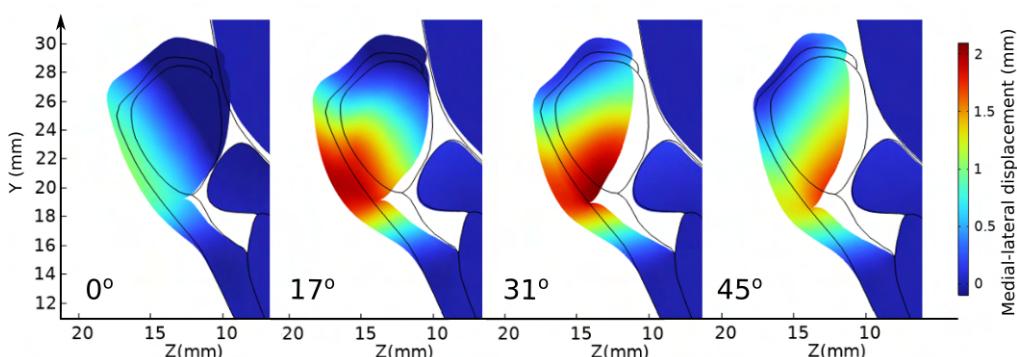


Fig. 3. Medial-lateral (z-axis) displacement distribution of the male larynx in a mid-membranous coronal plane for four different fiber angles. TA activation level was set to 1 (full activation). The solid line shows the initial resting position of the cover layer and TA muscle.

References:

- [1] Hirano, M. (1988). Vocal mechanisms in singing: laryngological and phoniatric aspects. *Journal of Voice*, 2(1), 51-69.
- [2] Zhang, Z. (2023). Vocal fold vertical thickness in human voice production and control: A review. *Journal of Voice*.
- [3] Vahabzadeh-Hagh, A. M., Zhang, Z., & Chhetri, D. K. (2017). Quantitative evaluation of the in vivo vocal fold medial surface shape. *Journal of Voice*, 31(4), 513-e15.
- [4] Yin, J., & Zhang, Z. (2016). Laryngeal muscular control of vocal fold posturing: Numerical modeling and experimental validation. *The Journal of the Acoustical Society of America*, 140(3), EL280-EL284.
- [5] Jiang, W., Geng, B., Zheng, X., & Xue, Q. (2024). A computational study of the influence of thyroarytenoid and cricothyroid muscle interaction on vocal fold dynamics in an MRI-based human laryngeal model. *Biomechanics and Modeling in Mechanobiology*, 1-13.
- [6] Zhang, Z., Samajder, H., & Long, J. L. (2017). Biaxial mechanical properties of human vocal fold cover under vocal fold elongation. *The Journal of the Acoustical Society of America*, 142(4), EL356-EL361.
- [7] Lehoux, S., & Zhang, Z. (2024). A Methodology to Quantify the Effective Vertical Thickness of Prephonatory Vocal Fold Medial Surface. *Journal of Voice*.

Engineered Composite Hydrogels for In Situ Tissue Repair: Enhanced Adhesion, Porosity, and Immunomodulation

Sara Nejati¹, Luc Mongeau¹

¹ Department of Mechanical Engineering, McGill University, Montreal, Quebec, Canada

Keywords: Hydrogel, Adhesive, Microporous

Abstract

Introduction & rationale: Injectable hydrogels can support the body's innate healing capability following defect or lesion removal by providing a temporary matrix for host cell ingrowth and regeneration. However, the clinical adoption of current injectable systems remains low due to issues like product dislodgment during administration, nanoporous structures that limit cell infiltration, and uncontrolled biological responses at the treatment site. Moreover, the removal of tissue defects and the subsequent implantation of tissue engineering scaffolds inevitably elicit injury, triggering inflammatory responses. While inflammation serves as a vital response for the body to initiate healing and safeguard itself, prolonged or excessive inflammation can lead to fibrosis, which is a pathological condition.

Objectives: This study aims to develop an injectable hydrogel optimized for in situ vocal fold (VF) tissue repair post-defect removal. The ideal hydrogel for this application should (i) cure rapidly to prevent leakage, (ii) adhere strongly to avoid dislodgment, (iii) possess mechanical properties conducive to VF regeneration, (iv) have a microporous structure to support cell infiltration and nutrient/oxygen transport, and (v) contain bioactive properties to promote regenerative healing and minimize fibrosis [1, 2].

Methods: We synthesized a double-network hydrogel by combining dopamine-grafted hyaluronic acid (DAHA) with silk fibroin (SF), which jellifies rapidly through Fe³⁺-dopamine coordination and is further strengthened by sonication-induced β-sheet formation in SF [2]. Curcumin-loaded polylactic acid (PLA) particles were incorporated to modulate the inflammatory response. The resultant composite hydrogel, designated as Microporous Double Network Composite (MDNC), comprises an optimized blend of curcumin-loaded PLA particles, 2.5% SF, 2.5% DAHA, and 10 mM iron ions. For comparative analysis, three additional formulations were prepared: Nanoporous Single Network (NSN) containing 5% SF, Microporous Single Network (MSN) composed of 5% DAHA, and Microporous Double Network (MDN) with equal parts of SF and DAHA.

Results: Scanning Electron Microscopy (SEM) and confocal microscopy of FITC-labeled hydrogels illustrated a highly porous structure of MDNC, with an average pore size of 110.21 μm. Rheological analysis revealed rapid gelation within 5 seconds due to iron-dopamine tris coordination, followed by a 50-minute stiffening phase due to β-sheet formation in SF. Oscillatory rheological tests showed that MDNC reached a storage modulus (G') of approximately 1000 Pa, suitable for VF repair, and exhibited a yield strain of about 410%. MDNC hydrogel also exhibited self-healing properties, recovering its original mechanical strength post-deformation. Young's modulus for the MDNC was around 5 kPa, compared to 21 kPa for NSN and 1.7 kPa for MSN. Adhesion tests revealed that MDNC hydrogel had a tensile adhesive strength of 35 kPa and outperformed commercial fibrin glue, which demonstrated a strength of 34 kPa. Biocompatibility was assessed through Live/Dead assays with over 90% cell viability observed. Phalloidin/DAPI staining confirmed healthy, fibroblast-like cell morphology. Cell migration studies showed a migration index of 19.3%. Anti-inflammatory and anti-fibrotic responses were highlighted by decreased α-SMA and COL1A1 expressions in TGF-β1 stimulated fibroblasts treated with curcumin, and enhanced CD206 expression in THP-1-derived macrophages by 2.9-fold, indicating a shift towards reparative M2 macrophage phenotypes [3].

Conclusions: The present study introduces a novel microporous double network composite (MDNC) hydrogel system, designed for effective in situ tissue repair following a defect or lesion removal. The hydrogel demonstrates adhesive strengths comparable to that of commercially available adhesives, such as fibrin glue. Their microporous structure facilitates rapid diffusion and effective cell penetration. Their double network configuration provides superior mechanical behavior. The sustained release of curcumin modulates the local immune response by promoting macrophage polarization towards the reparative M2 phenotype while counteracting the production of TGF-β1-induced collagen production and fibroblast differentiation into myofibroblasts, thereby inhibiting fibrosis. Thanks to an unprecedented combination of mechanical, structural, and biological properties, the proposed material is expected to impact broadly the repair and regeneration of mechanically dynamic tissues.

References:

- [1] Walimbe, T., Panitch, A., & Sivasankar, P. M. (2017). A review of hyaluronic acid and hyaluronic acid-based hydrogels for vocal fold tissue engineering. *Journal of Voice*, 31(4), 416-423.
- [2] Nejati, S., & Mongeau, L. (2023). Injectable, pore-forming, self-healing, and adhesive hyaluronan hydrogels for soft tissue engineering applications. *Scientific Reports*, 13(1), 14303.
- [3] Xu, Y., & Liu, L. (2017). Curcumin alleviates macrophage activation and lung inflammation induced by influenza virus infection through inhibiting the NF- κ B signaling pathway. *Influenza and other respiratory viruses*, 11(5), 457-463.

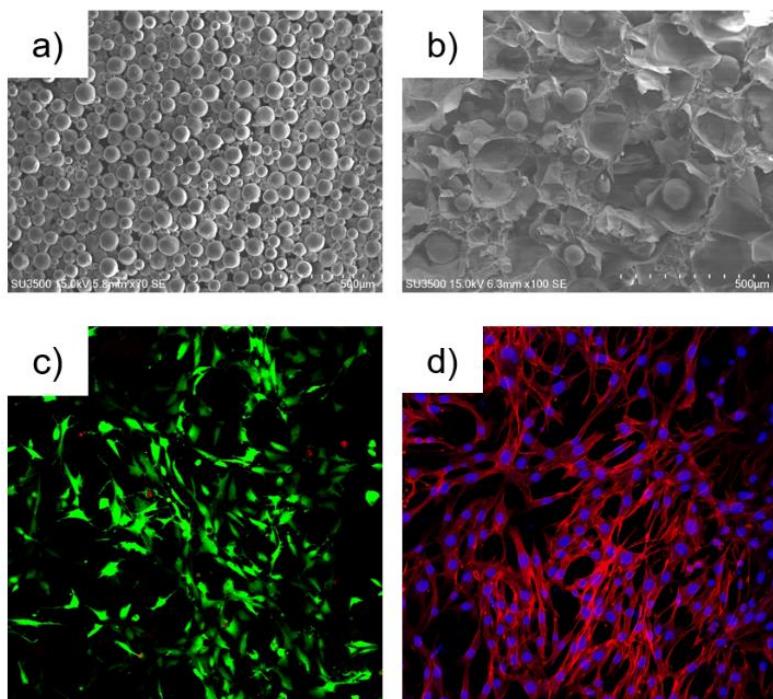


Figure 1. **a)** SEM image of curcumin-loaded PLA particles. **b)** SEM image of the MDNC hydrogel. **c)** live/dead assay on MDNC. **d)** Phalloidin/DAPI staining on MDNC hydrogel.

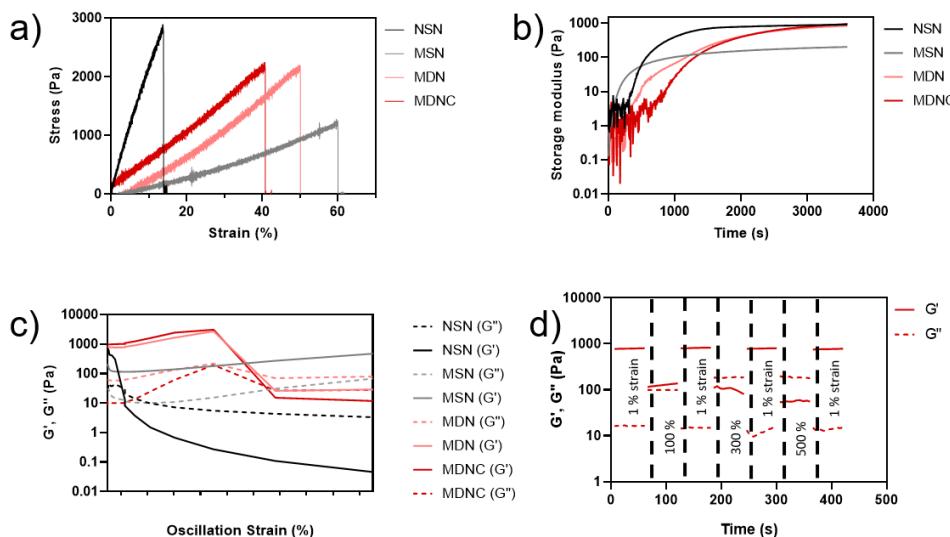


Figure 2: **a)** Stress-strain curves from tensile testing. **b)** Time-sweep test displaying gelation kinetics. **c)** Amplitude-sweep test illustrating yield strain. **d)** Oscillatory time sweeps demonstrating self-healing properties.

Psychosocial Stress-Induced Changes in Vocal Fold Epithelial Barrier Integrity

Anumitha Venkatraman¹, Katelyn Jacobs¹, Susan Thibeault¹

¹ Department of Surgery, Division of Otolaryngology Head and Neck Surgery, University of Wisconsin Madison, Madison, WI, USA

Keywords: vocal fold epithelium, psychosocial stress, immunofluorescence, epithelial barrier integrity

Abstract

Introduction & rationale: Epithelium is the first layer of defense in mucosal organs, limiting pathogenic access and infiltration.^{1,2} Following psychosocial stress exposure, other mucosal organs such as the gut exhibits increased epithelial permeability.¹⁻⁴ Structural and functional aspects of the stress-altered epithelia of the gut include decreased mucus thickness^{5,6}, downregulation of mucus genes,^{5,6} decreased epithelial tight and adherens junction integrity^{5,7,8} responsible for maintaining cytoskeletal structure and adherence to the basement membrane,² and upregulation of toll like receptors that recognize microbes on the epithelial surface and trigger downstream immune responses^{9,10}. Research indicates that a longer duration of stress exposure^{12,13} and females¹³ have worsened stress-induced epithelial barrier dysfunction, in the gut. However, the effect of psychosocial stress on vocal fold epithelial barrier integrity remains unknown. Increased vocal fold epithelial permeability can leave the larynx susceptible to the effects of noxious stimuli² and trigger inflammatory responses. Investigating stress-induced epithelial effects in the larynx could lay the groundwork for better understanding a role for stress in host-microbe immune interactions.

Objectives: To delineate changes in vocal fold epithelial barrier integrity following acute and chronic psychosocial stress. We hypothesized that acute and chronic psychosocial stress exposure would result in reduced vocal fold epithelial barrier integrity characterized by downregulation of mucins, adherens and tight junctions, upregulation of toll-like receptors and inflammatory cytokines. We hypothesized reduced vocal fold epithelial barrier integrity for chronic – compared to acute – psychosocial stress and females compared to their male counterparts.

Methods: C56BL/7 mice (N=42, equal males and females) were equally allocated into acute psychosocial stress, chronic psychosocial stress and control groups. Animals in the psychosocial stress groups were exposed to restraint stress for 6 hours a day. Acute psychosocial stress lasted for 7 days and chronic psychosocial stress lasted for 14 days with the control group undisturbed for the same duration. Corticosterone plasma levels were collected to confirm successful stress induction following acute and chronic psychosocial stress. Following the corresponding stress protocols, larynges were collected and prepared for downstream analysis. Immunofluorescence determined protein expression of epithelial tight and adherens junctions in coronal formalin-fixed paraffin-embedded 4 um sections of mid membranous vocal fold tissue. Quantitative polymerase chain reaction was used to compare relative gene expression of mucins, toll like receptors and inflammatory cytokines in vocal fold tissue across groups.

Results: Acute psychosocial stress resulted in downregulation of inflammatory cytokines (*IL 1β* (p = .018), *TNF α* (p = .012)) and mucins (*MUC2* (p = .047) compared to controls. Chronic psychosocial stress resulted in significant downregulation of (*IL 1β* (p = .006), *TNF α* (p = .026)) and mucin (*MUC2* (p = .047) compared to the controls. Toll like receptor 2 (TLR2) gene expression was significantly downregulated with chronic psychosocial stress only (p = .006), compared to controls. Visual inspection of immunofluorescence for proteins in epithelial adherens and tight junctions demonstrated in acute and chronic psychosocial stress – compared to control group – E cadherin in consistently observed throughout the mid-membranous portion of the vocal fold epithelium. Zo-1 was decreased in expression and inconsistent across the length of the mid-membranous vocal fold epithelium (greater in the inferior vocal fold epithelium compared to superior vocal fold epithelium) following acute and chronic psychosocial stress. There were no differences across the sexes across all outcomes.

Conclusions: Vocal fold epithelial barrier integrity is altered after acute and chronic psychosocial stress. There were no sex-related differences. Changes in vocal fold epithelial barrier integrity could leave the larynx susceptible to noxious environmental and systemic stimuli. These results lay the groundwork for delineating a role for psychosocial stress in the development of laryngeal pathologies.

References:

1. Gareau M, Silva M, Perdue M. Pathophysiological Mechanisms of Stress-Induced Intestinal Damage. CMM. 2008;8(4):274-281. doi:10.2174/156652408784533760
2. Levendoski EE, Leydon C, Thibeault SL. Vocal Fold Epithelial Barrier in Health and Injury: A Research Review. Journal of Speech, Language, and Hearing Research. 2014;57(5):1679-1691. doi:10.1044/2014_JSLHR-S-13-0283

3. Bailey MT, Engler H, Sheridan JF. Stress induces the translocation of cutaneous and gastrointestinal microflora to secondary lymphoid organs of C57BL/6 mice. *Journal of Neuroimmunology*. 2006;171(1- 2):29-37. doi:10.1016/j.jneuroim.2005.09.008
4. Everson CA, Toth LA. Systemic bacterial invasion induced by sleep deprivation. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*. 2000;278(4):R905-R916. doi:10.1152/ajpregu.2000.278.4.R905
5. Lin R, Wang Z, Cao J, Gao T, Dong Y, Chen Y. Role of melatonin in intestinal mucosal injury induced by restraint stress in mice. *Pharmaceutical Biology*. 2020;58(1):342-351. doi:10.1080/13880209.2020.1750659
6. Gao X, Cao Q, Cheng Y, et al. Chronic stress promotes colitis by disturbing the gut microbiota and triggering immune system response. *Proc Natl Acad Sci USA*. 2018;115(13). doi:10.1073/pnas.1720696115
7. Mazzon E, Cuzzocrea S. Role of TNF- α in ileum tight junction alteration in mouse model of restraint stress. *American Journal of Physiology-Gastrointestinal and Liver Physiology*. 2008;294(5):G1268-G1280. doi:10.1152/ajpgi.00014.2008
8. Machorro-Rojas N, Sainz-Espuñes T, Godínez-Victoria M, et al. Impact of chronic immobilization stress on parameters of colonic homeostasis in BALB/c mice. *Mol Med Report*. Published online June 27, 2019. doi:10.3892/mmr.2019.10437
9. Pandey S, Kawai T, Akira S. Microbial Sensing by Toll-Like Receptors and Intracellular Nucleic Acid Sensors. *Cold Spring Harb Perspect Biol*. 2015;7(1):a016246. doi:10.1101/cshperspect.a016246
10. Weber MD, Frank MG, Sobesky JL, Watkins LR, Maier SF. Blocking toll-like receptor 2 and 4 signaling during a stressor prevents stress-induced priming of neuroinflammatory responses to a subsequent immune challenge. *Brain, Behavior, and Immunity*. 2013;32:112- 121. doi:10.1016/j.bbi.2013.03.004
11. Venkatraman, Jacobs, Binns, An, Rey, Thibeault (in review). Effects of psychosocial stress on laryngeal microbiota and epithelial barrier integrity. *submitted to mSphere*
12. Weber MD, Frank MG, Sobesky JL, Watkins LR, Maier SF. Blocking toll-like receptor 2 and 4 signaling during a stressor prevents stress-induced priming of neuroinflammatory responses to a subsequent immune challenge. *Brain, Behavior, and Immunity*. 2013;32:112- 121. doi:10.1016/j.bbi.2013.03.004
13. Xu M, Wang C, Krolick KN, Shi H, Zhu J. Difference in post-stress recovery of the gut microbiome and its altered metabolism after chronic adolescent stress in rats. *Sci Rep*. 2020;10(1):3950. doi:10.1038/s41598-020-60862-1

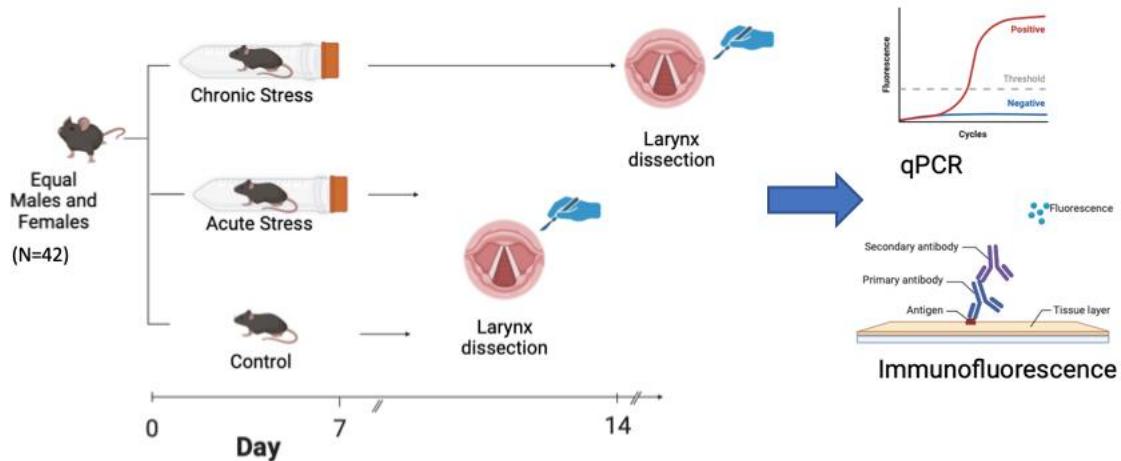


Figure 1: Experimental Protocol

PODIUM SESSION 4

Injectable Click Decellularized Extracellular Matrix Hydrogels for Vocal Fold Tissue Engineering: An *in vivo* Biocompatibility Study

Mika Brown¹, Hideaki Okuyama², Maryam Tabrizian^{1,3}, Nicole Y.K. Li-Jessen^{1,2,4}

¹Department of Biomedical Engineering, McGill University, Montreal, Quebec, Canada

² School of Communication Sciences and Disorders, McGill University, Montreal, Quebec, Canada

³ Faculty of Dental Medicine and Oral Health Sciences, McGill University, Montreal, Quebec, Canada

⁴ Department of Otolaryngology – Head and Neck Surgery, McGill University, Montreal, Quebec, Canada

Keywords: injectable hydrogels, decellularized extracellular matrix, vocal fold tissue engineering

Introduction & Rationale: **Decellularized extracellular matrix (dECM)** hydrogels have been used to develop injectable regenerative biomaterials for many organs. [1] dECM are derived from native tissues and thus retain essential innate cues for stimulating cellular recruitment and wound repair for tissue-specific regeneration. **Vocal fold-derived dECM (VF-dECM)** has low scalability due to the small size of the tissue. Meanwhile, hydrogels derived from **small intestinal submucosa-derived dECM (SIS-dECM)** have shown promise in vocal fold repair. [2,3]

In our previous work, we investigated the importance of tissue-specificity in VF repair by comparing the proteomic composition of porcine SIS- and VF- dECM. [2] We found that SIS-dECM possessed lower in-group variability in proteomic composition. Additionally, SIS- and VF- dECM hydrogels stimulated the production of VF-relevant neo-ECM components, elastin and hyaluronan, by human vocal fold fibroblasts at comparable levels, significantly greater than collagen hydrogel controls.

Meanwhile, dECM hydrogels often suffer from rapid and variable degradation rates *in vivo*. We hypothesized that using bioorthogonal click chemistry to produce a composite dECM hydrogel with supportive **alginate (Alg)** linkages would improve the longevity of dECM hydrogels while retaining biocompatibility for use in vocal fold augmentation.

Objectives: In this study, we aimed to (1) investigate the biocompatibility and immune response to **dECM-alginate (dECM-Alg)** hydrogels crosslinked by click chemistry in a subcutaneous rat model and (2) compare the immune response to hydrogels made from SIS- and VF- dECM-Alg hydrogels respectively.

Methods: Porcine VF were decellularized using our established protocol, with a series of incubation steps in hypertonic salt, nucleases, and isopropanol.[2] Homogenization was performed using a freezer mill. SIS dECM was generously donated by Cook Biotech as a powder. dECM powders were solubilized with pepsin and their primary amines functionalized by carbodiimide reaction with NHS-conjugated norbornene (Nb). dECM-Nb was purified by dialysis and lyophilized product was sterilized using ethylene oxide. Alginate (Alg) was functionalized with amine-conjugated methyltetrazine (Tz) by carbodiimide reaction and purified by dialysis and sterile filtration. Click dECM-Alg hydrogels were fabricated by mixing pepsin solubilized dECM-Nb with Alg-Tz at a 4:1 ratio.

The animal study was approved by the McGill University Animal Compliance Office (Protocol MCGL-8275). Two dECM-Alg derived injectables, namely **SIS-Alg** and **VF-Alg**, were tested for their biocompatibility. A total of 10 Sprague-Dawley rats, 5-20 weeks old (100-500 g) were anesthetized with isofluorane and subcutaneously injected with a 21-gauge needle with 200 µL SIS-Alg, VF-Alg pre-gels or PBS as controls. At Day 7 and Day 21, animals were euthanized (N=5 for each time point) and their skin samples were excised for histological evaluation.

Hematoxylin and eosin staining (H&E) and immunofluorescence were performed on 10 µm thick cryosections. For immunofluorescence staining, skin sections were permeabilized and incubated with rabbit anti-CD68 (1:500 dilution, Invitrogen PA5-78996) and rabbit anti-CD206 (1:200 dilution, Proteintech 18704-1) to detect total macrophages and M2 polarized macrophages respectively. The next day, sections were incubated for 1 h with DAPI and secondary antibodies, Alexa Fluor® 488 or Alexa Fluor® 647 Goat anti-rabbit IgG (1:1000) followed by mounting with Fluoromount-G™. Both brightfield and fluorescent imaging was performed on a Zeiss Axio Observer 3, and analysis performed using ImageJ. For immunofluorescent samples, the number of CD68+ and CD206+ cells in the interior and immediate exterior of the hydrogels was evaluated to identify cellular infiltration within the hydrogels and their impact on surrounding tissue.

Results: H&E stains revealed that hydrogels remained visible in all rats after day 7 and 21, except one SIS-Alg animal at day 21. This sample was excluded from the subsequent analysis as the gel area could not be identified under the skin. A strong immune response was observed on day 7 for both gels, as indicated by a dense fibrous capsule around

the hydrogels. On day 21, the dense cell border was absent, while cells accumulated near the center of the gel area for all click dECM hydrogel samples. (**Figure 1**) Robust clustering of cells expressing CD68, a pan-macrophage marker, was observed around the edges and within SIS-Alg hydrogels on Day 7 (interior: 1410 ± 896 , exterior: 2786 ± 1262) (**Figure 2A**). A significant decrease in the density of CD68+ macrophages in the exterior of SIS-Alg hydrogels (916 ± 362) was observed on Day 21. A proportion of macrophages expressing the pro-reconstructive marker CD206 were found on the exterior and interior of SIS-Alg on day 7 (interior: 716 ± 402 , exterior: 1386 ± 247) (**Figure 2B**). In the exterior of the SIS-Alg hydrogels, a decrease in CD206+ macrophages was observed from Day 7 to Day 21 (530 ± 199). The average ratio of CD206+/CD68+ macrophages was 0.6 or greater for the exterior and interior of SIS-Alg on both day 7 and 21, meaning the majority of macrophages exhibited M2 polarization (**Figure 2C**).

For VF-Alg hydrogels, CD68+ (interior: 1967 ± 787 , exterior: 2750 ± 809) and CD206+ ((interior: 732 ± 421 , exterior: 1097 ± 300) macrophages were likewise found in the hydrogel interior and around the hydrogel edges on Day 7. On Day 21, the density of CD68+ macrophages on the interior and exterior of VF-Alg hydrogels (interior: 881 ± 93 , exterior: 857 ± 307) significantly decreased from Day

7. A significant decrease in CD206+ macrophage density was also observed in the exterior of VF-Alg hydrogels on Day 21 (351 ± 154). The average ratio of CD206+/CD68+ macrophages on the interior and exterior of VF-Alg hydrogels was below point 0.5 for both day 7 and day 21.

In comparison to the PBS controls, VF-Alg and SIS-Alg hydrogels showed higher density of CD68+ and CD206+ macrophages in both exterior and interior regions at all time points, suggesting that the macrophage response remained elevated at the end point of the experiment.

Conclusions: This study confirmed the biocompatibility of the click dECM hydrogels through the integration of the hydrogels with surrounding tissue. While the pan macrophage population remained elevated in our subcutaneous rat skin samples after 21 days, a significant decrease was observed in the exterior of the hydrogels. Notably, the higher ratio of CD206+ M2 polarized macrophages implicates a possible favorable anti-inflammatory and regenerative macrophage response to hydrogels produced from SIS compared to VF dECM. Future study includes the validation of click SIS dECM-Alg hydrogels in larger animals for longer term VF-specific regeneration.

Funding Support: This study was supported by the National Sciences and Engineering Research Council of Canada (RGPIN-2018-03843, RGPIN-2024-04235 and ALLRP 548623-19), Canada Research Chair research stipend (J.Y.L., M.T. and N.Y.K.L.-J.) and the National Institutes of Health (R01 DC-018577-01A1). The presented content is solely the responsibility of the authors and does not necessarily represent the official views of the above funding agencies

References:

- [1] Brown, M., Okuyama, H., Yamashita, M., Tabrizian, M., & Li-Jessen, N. Y. (2024). Trends in Injectable Biomaterials for Vocal Fold Regeneration and Long-Term Augmentation. *Tissue Engineering Part B: Reviews*.
- [2] Brown, M., Zhu, S., Taylor, L., Tabrizian, M., & Li-Jessen, N. Y. (2023). Unraveling the Relevance of Tissue-Specific Decellularized Extracellular Matrix Hydrogels for Vocal Fold Regenerative Biomaterials: A Comprehensive Proteomic and In Vitro Study. *Advanced nanobiomed research*, 3(4), 2200095.
- [3] Hu, J. J., Lei, X. X., Jiang, Y. L., Zou, C. Y., Song, Y. T., Wu, C. Y., ... & Xie, H. Q. (2022). Scarless vocal fold regeneration by urine-derived stem cells and small intestinal submucosa hydrogel composites through enhancement of M2 macrophage Polarization, neovascularization and Re-epithelialization. *Smart Materials in Medicine*, 3, 339-351.

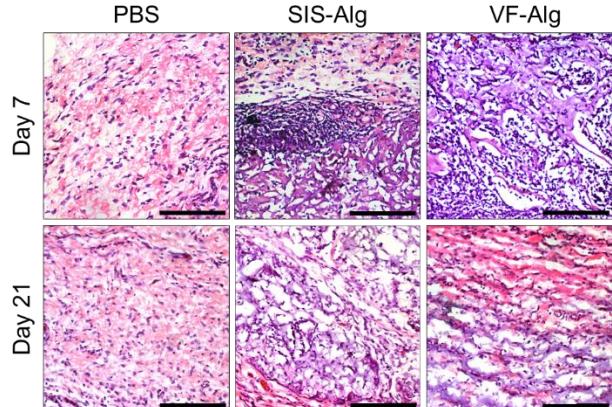


Figure 1. Representative H&E stain of subcutaneous rat tissue injected with click dECM-Alg hydrogels or PBS. Click dECM hydrogels stained purple due to the presence of alginate.

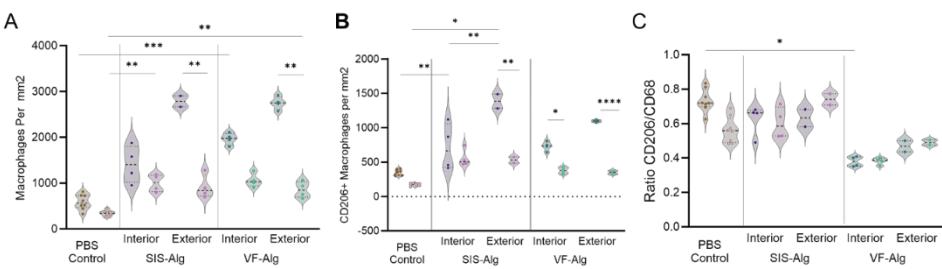


Figure 2. Macrophage infiltration and polarization in subcutaneous rat tissue injected with click dECM hydrogels or PBS. **A)** Total macrophages (CD68+) per square millimeter of tissue. **B)** CD206+ macrophages per square millimeter of tissue. **C)** Ratio of CD206+ macrophages to total CD68+ macrophages. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$. Error bars represent standard deviations.

A Multi-Material Hydrogel Model to Study Chronic Fibrosis in Vocal Folds

Mohaddeseh Mohammadi¹, Ayda Pourmostafa¹, Daniel Li¹, Hossein Goodarzi², Amir K. Miri^{1,3}

¹ Department of Biomedical Engineering, New Jersey Institute of Technology (NJIT), Newark, NJ, USA

² University Medical Center Göttingen, Robert-Koch-Str. 40, 37075, Göttingen, Germany

³ Department of Mechanical and Industrial Engineering, New Jersey Institute of Technology, Newark, NJ, USA

Keywords: vocal folds; fibroblasts; extracellular matrix; layer-wise.

Abstract

Vocal fold (VF) fibrosis is known as excessive collagen deposition in the extracellular matrix (ECM), resulting in tissue stiffness, disrupted vibration, and permanent impairment of the voice production [1]. VF fibrosis can stem from factors such as intubations, radiations, and vocal abuse [2]. The VF scarring induces significant mechanical alterations that impair native tissue function. This is associated with inflammatory gene expression, cell proliferation, and differentiation. This study proposes a novel hydrogel bioreactor model to mimic VF mechanical properties and assess fibrotic markers under loading conditions for screening steroid-based therapeutic interventions. We applied hydrogel engineering and bioprinting methods to create a multi-material mimetic model. We employed human VF fibroblasts for our chronic fibrotic modeling. The platform can advance the understanding of VF fibrosis and provide a robust tool for screening therapies to prevent fibrosis-related atrophy and restore functionality.

Introduction & rationale:

Approximately 9% of the general population has a voice abnormality during their lifespan [1-3]. One of the severe conditions affecting voice is when part of the soft and pliable mucosa is lost or replaced by dense fibrous tissue, leading to chronic VF fibrosis [4, 5]. VF fibrosis can be congenital or acquired following phono-trauma or chronic inflammatory pathologies [6]. Chronic VF fibrosis can be treated locally using a subepithelial injection of VF biomaterials or biological agents [7-9]. To address this need, we aim to mimic the lamina propria of the VF, which can be divided into three subsections (Superficial, Intermediate, and Deep). A multi-layered hydrogel was developed to account for these sections with similar mechanical properties for each layer. Gelatin Methacrylate was used as our hydrogel platform, as it is a form of denatured collagen that can behave biologically similarly to the ECM of the lamina propria.

Objectives:

This study aims to develop a multi-layered hydrogel bioreactor model that replicates the mechanical properties of the VF (see **Figure 1**). Using Gelatin Methacrylate, the model provides a physiologically relevant platform to study VF fibrosis, assess fibrotic markers under mechano-transduction conditions, and evaluate potential therapies to restore VF functionality.

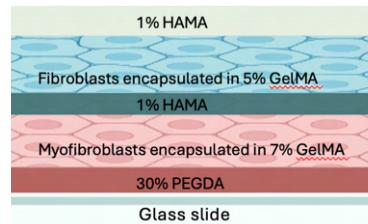


Figure 1. The model description.

Methods:

We developed a layer-wised hydrogel bioreactor to replicate the mechanics of VF tissue, including stiffness gradients and porosity, to mimic ECM conditions observed in both healthy and fibrotic VFs. Our protocol optimized the preparation of gelatin methacryloyl (GelMA; 3, 5, and 7% w/v) and hyaluronic acid methacrylate (HAMA; 0.5 and 1% w/v) hydrogels, achieving stiffness values within the range typical for fibrotic VF tissue (6-10 kPa). We used digital light processing (DLP) bioprinting to create layer-wise structures, as established by our group. We performed standard compression testing, shear rheometry, and degradation testing to characterize the material model (see **Figure 2**). Human-derived VF fibroblasts (HVOX; with the courtesy of Prof. Ryan Branski, New York University, NY), encapsulated in hydrogels at concentrations of 50×10^3 and 500×10^3 cells/mL, were photopolymerized with a hydrogel model for structural stability (following our preliminary observations). The selected biomarker TGF- β 1, indicating fibroblast to myofibroblast differentiation, was reconstituted in 10 Citric Acid, pH 3.0 to 0.1-1.0 mg/ml. After initial reconstitution, it was further diluted in a buffer containing 0.1%BSA. Activin was dissolved in 1 mg/ml BSA in phosphate-buffered saline and used at 10 to 20 ng/ml concentrations. After addition, cells were further incubated for 12 to 72 hours. The hydrogels were characterized, and physical and mechanical properties were monitored over one week to assess consistency with VF tissue (not shown here). Cell viability was tracked using a metabolic activity assay (CCK8, Sima-Aldrich), while cell phenotypes and behavior were analyzed via immunostaining for polymerized Phalloidin, F-actin, and DAPI at different time points (following standard assays; Life Technologies Corp.).

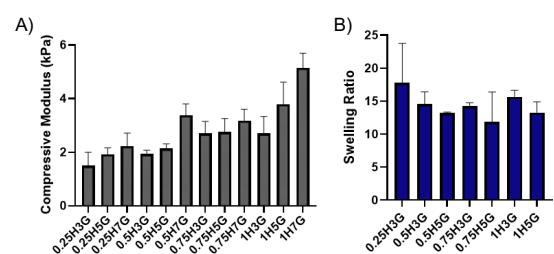


Figure 2. Selected physical properties of our model: A) compressive modulus (kPa) of our biomaterial groups, B) swelling ratio of our groups. (xHyG: x% of HAMA and y% of GelMA)

Results:

The proposed material model can mimic the elastic properties of VF tissue according to our physical characterization. The gel stability over one week confirmed the application of the model for potential chronic fibrosis modeling. The metabolic activity of the hydrogel groups indicates the negative impact of a high concentration of HAMA compared to GelMA. This can be related to a lack of cell moieties in the hydrogel. Our data show that VF fibrosis involves upregulated biomarkers such as TGF- β 1 and COL11A1 and increased collagen I deposition, contributing to stiffness and impaired elasticity (not shown here). The bioreactor model replicates these features, providing valuable insights into cellular responses to treatments. Live-dead cell assay revealed that HVOX cells exhibited high viability and greater spreading in softer hydrogels, such as 0.5% HAMA and 3% GelMA (i.e., 0.5H3G), compared to stiffer hydrogel (Figure 3A-B). Metabolic activity peaked at Day 3 in softer hydrogels. It declined progressively in all high-ratio HAMA compositions (Figure 3C).

In contrast, stiffer matrices impose mechanical constraints that inhibit these processes. Furthermore, using this hydrogel model, we studied how ECM stiffness influences fibroblast branching, with stiffer environments leading to reduced branching, which may offer a further understanding of fibrosis progression and potential therapeutic strategies. To confirm our TGF- β protocol, we cultured two replicates of treated and untreated cells in a 24-well plate with a cell density of 20×10^3 cells/ml (data not shown here). After 24 hours, the cells were stained using Phalloidin, F-actin, and DAPI for further verification.

Conclusions:

This study presents a novel bioreactor replicating the mechanical and biological environment of fibrotic VF tissue, enabling more precise *in vitro* analysis of fibrotic responses. Closely mimics the mechanical and biological environment of fibrotic VF tissue, supporting more accurate *in vitro* analysis of fibrotic responses. The model provides a platform that can evaluate antifibrotic therapies and improve treatments for VF fibrosis, potentially preventing dysphonia. Future integration of TGF- β signaling pathways and controlled mechanical vibration within this model further elucidate their roles in fibrosis modulation. Furthermore, the impact of TGF- β signaling pathways and mechanical vibration within this hydrogel model will further explain and be explored in the context of our hydrogel model to better understand their roles in fibrosis modulation, paving the way for novel therapies to mitigate VF scarring and dysfunction [8-9].

References:

- Li, L., et al., Tissue engineering-based therapeutic strategies for vocal fold repair and regeneration. *Biomaterials*, 2016. 108: p. 91-110.
- Bhattacharya, N., The prevalence of voice problems among adults in the United States. *Laryngoscope*, 2014. 124(10): p. 2359-62.
- Roy, N., et al., Voice disorders in the general population: prevalence, risk factors, and occupational impact. *Laryngoscope*, 2005. 115(11): p. 1988-95.
- Nakamura, R., et al., Concurrent YAP/TAZ and SMAD signaling mediate vocal fold fibrosis. *Sci Rep*, 2021. 11(1): p. 13484.
- Xu, H. and G.-K. Fan, The Role of Cytokines in Modulating Vocal Fold Fibrosis: A Contemporary Review. *The Laryngoscope*, 2021. 131(1): p. 139-145.
- Velier, M., et al., Paracrine Effects Of Adipose-Derived Stromal/Stem Cells And Stromal Vascular Fraction In An In Vitro Fibrogenesis Model Of Human Vocal Fold Scarring. 2020.
- Hess, M.M. and S. Fleischer, Laryngeal framework surgery: current strategies. *Curr Opin Otolaryngol Head Neck Surg*, 2016. 24(6): p. 505-509.
- Benninger, M.S., et al., Vocal Fold Scarring: Current Concepts and Management. *Otolaryngology—Head and Neck Surgery*, 1996. 115(5): p. 474-482.
- Ford, C.N., Advances and refinements in phonosurgery. *Laryngoscope*, 1999. 109(12): p. 1891-900.

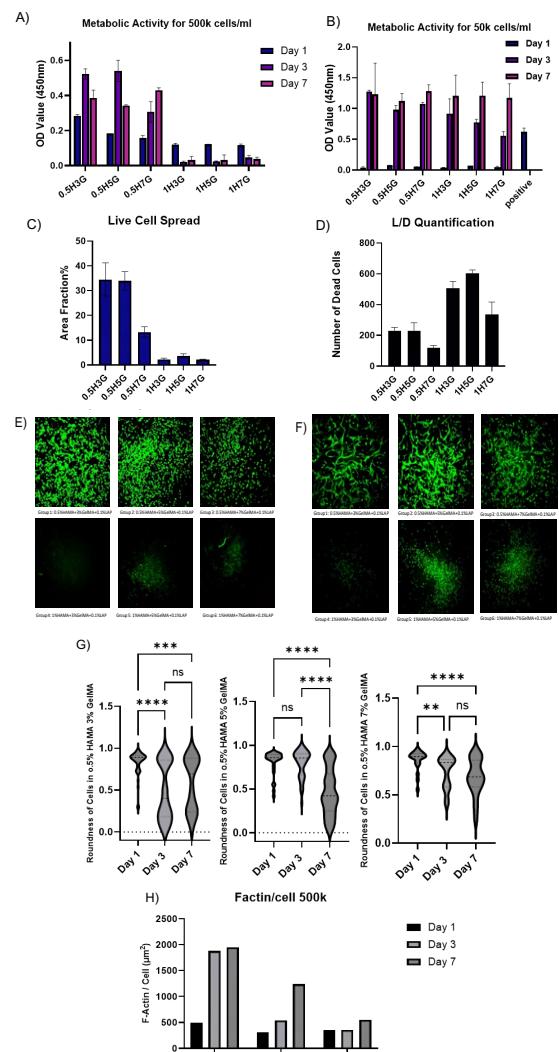


Figure 3. Biological characterization of our hydrogel model for two cell densities: A) metabolic activity for six groups measured, at Day 1, 3, and 7, for 500×10^3 cell/ml; B) 50×10^3 cell/ml; C) live cell spread quantification at Day 7; D) the number of dead cells for different groups at Day 7; E) live/dead fluorescent images for 500×10^3 cells/ml at Day 1 (green: live & red:dead); F) at Day 3; G) cell roundness for the hydrogel; E) F-actin is expressed per cell in soft & stiff samples at Day 21; F) cell spread area.

Innovating Treatments for Vocal Fold Scarring: Controlled Dexamethasone Delivery with a Novel Light-Activatable Implant

Denzel Ryan Cruz¹, Nour Awad², Avery Zheng³, Alexander Karabachev², Charles Farbos de Luzan², Yoonjee C. Park^{1,3}, Gregory R. Dion^{1,2}

¹ Department of Biomedical Engineering, University of Cincinnati College of Engineering and Applied Science, Cincinnati, Ohio, USA

² Department of Otolaryngology-Head and Neck Surgery, University of Cincinnati College of Medicine, Cincinnati, Ohio USA

³ Department of Chemical and Environmental Engineering, University of Cincinnati College of Engineering and Applied Science, Cincinnati, Ohio, USA

Keywords: Controllable Drug Delivery, Vocal Fold Scar, Vocal Fold Therapy

Abstract

Introduction & rationale: Voice disorders are the most common communication disorder in the US (Ramig, 1998). Vocal fold (VF) scarring, a common etiology for dysphonia, can occur following intubation-related trauma, post-radiation therapy, upper respiratory tract infection, voice misuse, smoking, and aging. Local irritation or injury along the delicate laryngotracheal structures can trigger inflammation and fibrosis (Branski, 2006). Depending on the nature of the lesion, surgical intervention is considered, but laryngeal surgery can worsen scarring with no effective replacement or regenerative treatment. Clinically, primary or adjuvant medical treatments like intralesional glucocorticoid injections are commonly employed to improve wound healing by minimizing fibrosis (Wang, 2013). Despite their ability to reduce scar formation, these effects may be short lived due to their rapid clearance and short half-lives, potentially limiting their long-term effectiveness (Wang, 2017). At present, the only options for repeat dosing involve in-office procedures or repeat surgery because access to the vocal folds for intralesional injection is challenging. Currently, there are no delivery systems capable of drug delivery over time to the VFs (Cruz, 2024).

Objectives: We have developed a biodegradable implant capable of dose-controllable delivery of glucocorticoids via laser light activation (Zheng, under review). This study assesses the efficacy of our light-activatable implant in a VF injury model to a single glucocorticoid injection. We hypothesize that our light-activatable implant will yield lasting wound healing effects and improved VF functionality in comparison a single-dose dexamethasone (Dex) injection, with minimal use of glucocorticoid.

Methods: 10 anesthetized New Zealand white rabbits underwent endoscopic VF injury followed by either (1) Dex injection or (2) implantation of a light-activatable Dex-loaded polymer capsule, with five injured and five uninjured VF controls. The single-dose Dex injection and implantation of the implant were administered anteriorly through the neck, where the needle was inserted through the cricothyroid membrane to treat the VF lesion. On days 0 and 21, VF implants were irradiated with a near-infrared pulsed laser (1064 nm, 100 mW) for 1 minute. Endoscopic imaging and follow-up evaluations were conducted on the same days. The larynges were harvested after 42 days to undergo ex vivo high-speed imaging and mechanical testing, followed by preparation for histological examination. 10 vibratory cycles were analyzed ex vivo using high-speed videos at >10 kHz to generate kymograms. For biomechanical measurements, an indentation map with at least 40 positions was oriented for each specimen, where normal force and structural stiffness and displacement at 1.96 mN were acquired. The mid-coronal zone along the VF edge was identified as the primary region of injury for data analysis as established previously by our group (Dion, 2019). Multivariate analysis of variance was conducted using Pillai's trace followed by post-hoc pairwise comparisons using Bonferroni correction to identify significant group differences. Histological staining was performed using hematoxylin and eosin to visualize VF structures and Masson's trichrome to identify collagen distribution. Each sample was immunohistochemically stained for Type I collagen and fibronectin to assess extracellular matrix composition.

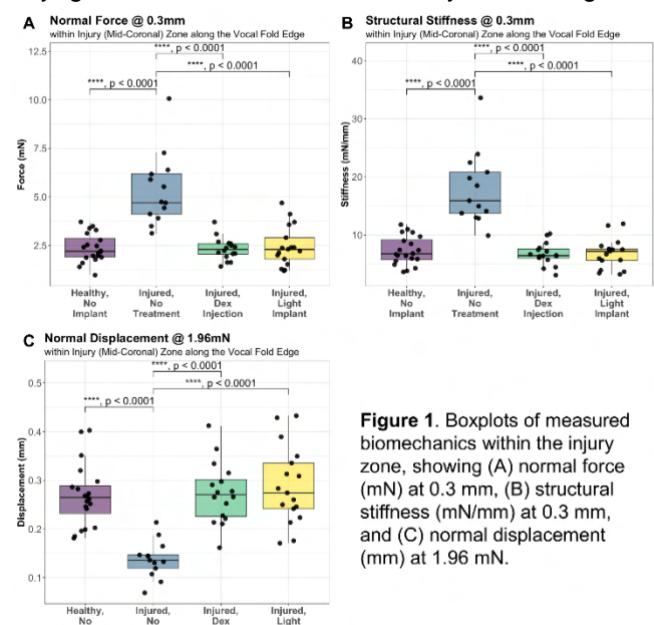


Figure 1. Boxplots of measured biomechanics within the injury zone, showing (A) normal force (mN) at 0.3 mm, (B) structural stiffness (mN/mm) at 0.3 mm, and (C) normal displacement (mm) at 1.96 mN.

Results: Biomechanical analysis of all rabbit VFs within the injury zone are summarized in Figure 1. Untreated/injured VFs exhibited significant changes across each dependent variable after 42 days in comparison to healthy VFs, with increases in normal force ($2.37 \pm .71$ mN vs. 5.37 ± 1.81 mN, $p < 0.0001$) and structural stiffness (7.31 ± 2.30 mN/mm vs. 17.97 ± 6.03 mN/mm, $p < 0.0001$) and a decrease in normal displacement ($.27 \pm .07$ mm vs. $.14 \pm .04$ mm, $p < 0.0001$). Compared to untreated/injured VFs, those with the implant markedly reduced normal force (5.37 ± 1.81 mN vs. 2.46 ± 1.00 mN, $p < 0.0001$) and structural stiffness (17.97 ± 6.03 mN/mm vs. 6.81 ± 2.46 mN/mm, $p < 0.0001$), and increased displacement at 1.96 mN ($.14 \pm .04$ mm vs. $.29 \pm .08$ mm, $p < 0.0001$). Injured VFs with a single-dose glucocorticoid also showed decreases in normal force (5.37 ± 1.81 mN vs. $2.31 \pm .56$ mN, $p < 0.0001$) and structural stiffness (17.97 ± 6.03 mN/mm vs. 6.70 ± 1.85 mN/mm, $p < 0.0001$) and an increase in normal displacement ($.14 \pm .04$ mm vs. $.27 \pm .06$ mm, $p < 0.0001$) in comparison to the injured control VFs. No differences in biomechanical outcomes were observed between implant and single-dose injection groups.

Heatmaps of the three biomechanical variables that include all data points and the designated injury zone for each specimen were created. The color scale was normalized to cover the entire range across all specimens, with red specifying increased stiffness and blue indicating decreased stiffness. Qualitatively, along the VF edge, higher values of structural stiffness (Figure 2, left column) extend farther out anteriorly in the injury control as seen with the presence of yellow in the injury zone. In the two treatment groups, there were no distinguishable marks to identify sites of injection or implantation within the borders of biomechanical testing, which covers the glottic and subglottic regions. 10 vibratory cycles from ex vivo high-speed videos were converted as kymograms to visualize VF mucosal wave oscillations within the designated injury zone along the VF edge (Figure 2, right column). The long ratio of time in the closed phase in the injury VF is shortened in both treatment groups. Images from histology and immunostaining were compiled for each specimen and compared between groups.

Conclusions: Comparable to a single-dose glucocorticoid injection, our light-activatable implant with dose-controllable delivery of glucocorticoid improves the VF biomechanical properties with changes in vibratory function and histology, supporting its therapeutic potential. Optimizing the dosage and activation timing of the implant may be warranted to further enhance the VF healing response.

References:

- Branski, R. C., Verdolini, K., Sandulache, V., Rosen, C. A., & Hebda, P. A. (2006). Vocal Fold Wound Healing: A Review for Clinicians. *Journal of Voice*, 20(3), 432–442. <https://doi.org/10.1016/j.jvoice.2005.08.005>
- Cruz, D. R. D., Zheng, A., Debele, T., Larson, P., Dion, G. R., & Park, Y. C. (2024). Drug delivery systems for wound healing treatment of upper airway injury. *Expert Opinion on Drug Delivery*, 21(4), 573–591. <https://doi.org/10.1080/17425247.2024.2340653>
- Dion, G. R., Guda, T., Mukudai, S., Bing, R., Lavoie, J., & Branski, R. C. (2019). Quantifying vocal fold wound-healing biomechanical property changes. *The Laryngoscope*, 130(2), 454–459. <https://doi.org/10.1002/lary.27999>
- Ramig, L. O., & Verdolini, K. (1998). Treatment Efficacy. *Journal of Speech, Language, and Hearing Research*, 41(1). <https://doi.org/10.1044/jslhr.4101.s101>
- Wang, C.-T., Lai, M.-S., & Cheng, P.-W. (2017). Long-term Surveillance Following Intralesional Steroid Injection for Benign Vocal Fold Lesions. *JAMA Otolaryngology–Head & Neck Surgery*, 143(6), 589. <https://doi.org/10.1001/jamaoto.2016.4418>
- Wang, C.-T., Lai, M.-S., Liao, L.-J., Lo, W.-C., & Cheng, P.-W. (2013). Transnasal endoscopic steroid injection: A practical and effective alternative treatment for benign vocal fold disorders. *The Laryngoscope*, 123(6), 1464–1468. <https://doi.org/10.1002/lary.23715>
- Zheng, A., Awad, N., Cruz, D. R. D., Pissay, R., Dion, G. R., & Park, Y. C. (Under Review). Controlled-Release of Dexamethasone Via Light-Activated Implant for Potential Vocal Fold Scar Treatment. *ACS Biomaterials Science & Engineering*.

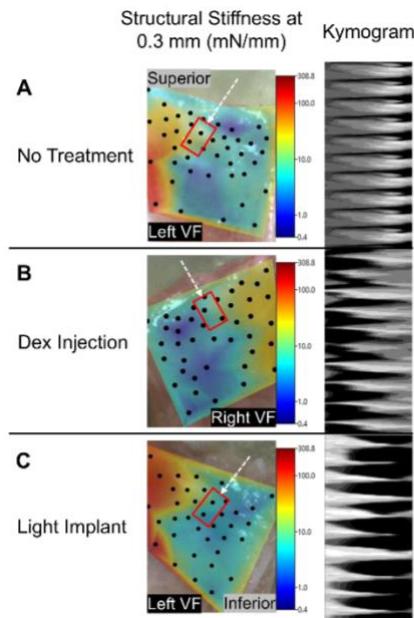


Figure 2. Representative heatmaps of structural stiffness and kymograms from an injured VF left (A) untreated or treated with (B) a single-dose injection or (C) a light-activatable implant. Heatmaps are oriented from superior (top) to inferior (bottom). Kymograms display 10 VF vibratory cycles. The red box on the heatmap indicates the injury zone (mid-coronal zone along VF edge), and the white dashed arrow is the approximate camera viewpoint during high-speed imaging to generate kymograms.

POSTER SESSION 1

Effects of Surgery on the Relationship between Subglottic Pressure and Fundamental Frequency in Vocal Fold Dynamics in Patients with Benign Laryngeal Diseases

Wen-Hsuan Tseng^{1,2}, Hsiang-Ling Chiu¹, Tzu-Yu Hsiao¹, Tsung-Lin Yang^{1,2,3*}

¹ Department of Otolaryngology, National Taiwan University Hospital and National Taiwan University College of Medicine, Taipei, Taiwan

² Graduate Institute of Clinical Medicine, National Taiwan University College of Medicine, Taipei, Taiwan

³ Research Center for Developmental Biology and Regenerative Medicine, National Taiwan University, Taipei, Taiwan

Keywords: *Fundamental frequency- Subglottic pressure- Treatment effect*

Abstract

Introduction & rationale:

Subglottic pressure (P_s) and fundamental frequency (F_0) play important roles in governing vocal fold (VF) dynamics. Theoretical description, model simulation, excised larynx and animal models have been used in previous studies, yet clinically applicable measurements are still lacking. The slope of the stress-strain curve of the VFs demonstrates the extent to which the VFs resist deformation in response to an applied force, indicating stiffness. It has been demonstrated in an *in vivo* canine model that F_0 increased linearly as P_s increased, but with a lesser slope for higher levels of VF tension. Kitajima et al. conducted *in vivo* testing of the relationship between F_0 and P_s for VF stiffness in human. They concluded that there is a relational pattern between the change in F_0 per unit change of P_s ($\frac{dF_0}{dP_s}$) and F_0 , having potential applications in pre- and post-treatment comparisons. Similarly, Scherer reported that the relationship between F_0 and P_s in participants with abnormally stiff VFs differs from that in healthy participants. In our previous study, the relationship between F_0 and P_s was investigated using the airflow interruption method in awake patients noninvasively without the need of anesthesia. And it had been shown that this relation, described as the slope (Hz/kPa) of the regression line of the frequency-pressure data pairs, decreased significantly as VF stiffness increased in healthy volunteers and that it could potentially be used as an analog for VF stiffness.

Objectives:

In the present study, we explored the applicability of the proposed method in patients with benign VF lesions, and aimed to evaluate the effects of surgery for benign VF lesions by investigating the relationship between F_0 and P_s .

Methods:

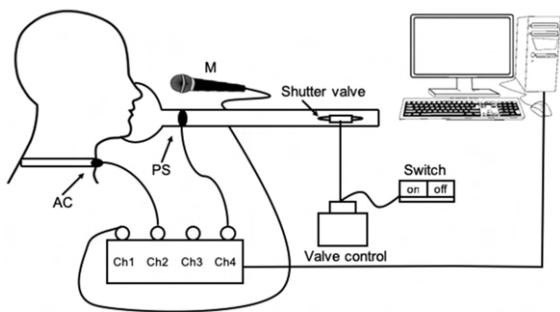
Patients with benign VF lesions who underwent phonosurgery were prospectively recruited. Participants were instructed to sustain voicing the vowel /o/ at three incremental frequencies four semitones apart in the modal register (F01, F02, F03). F_0 was estimated by VF vibration on the accelerometer. P_s change was achieved and measured using the airflow interruption method. Clinical voice outcomes and phonation threshold pressure (PTP) were also evaluated at pre- and post-operation for comparison.

Results:

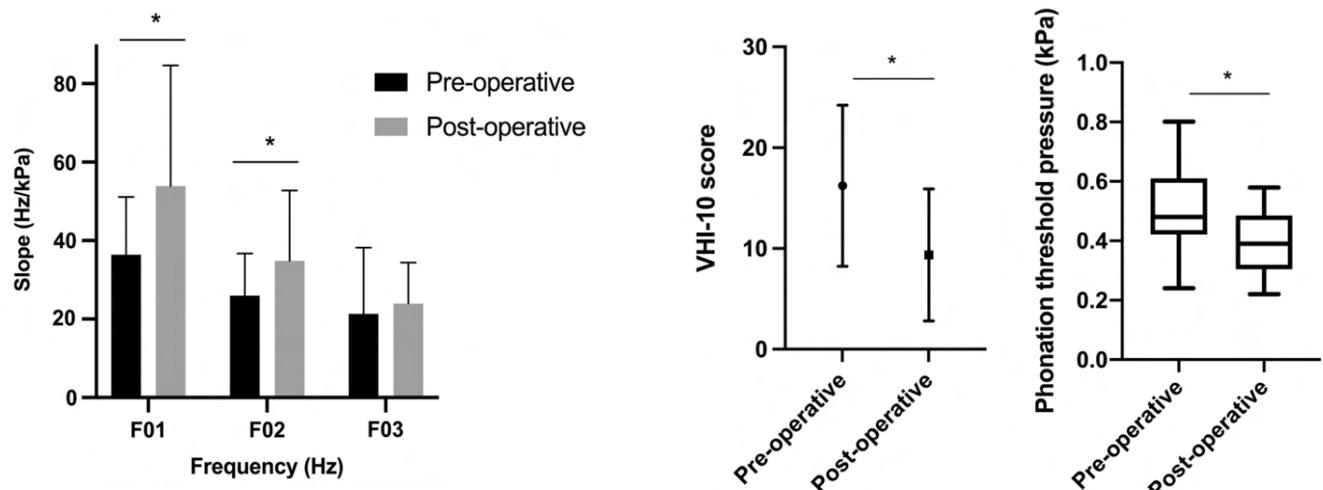
Thirteen patients with a mean age (SD) of 43.5 (12.4) years were included. The change in F_0 per unit change of P_s , which is the slope (Hz/kPa) of the regression line of the frequency-pressure data pairs, decreased as the tension of the VF increased. The slopes significantly increased after the operation for F01 and F02 (36.43 ± 14.68 preoperatively, 53.91 ± 30.71 postoperatively, $P=0.011$ and 26.02 ± 10.71 ; 34.85 ± 17.92 , $P = 0.046$, respectively). Additionally, there was a significant decrease in PTP and improvements in the grade, roughness, breathiness, asthenia, strain scale, and the voice handicap inventory-10.

Conclusions:

The relationship between F_0 and P_s may serve as an objective assessment of the outcomes in the treatment of benign laryngeal diseases with clinical relevance.



Schematic diagram of the airflow interruption device used in this study.
AC indicates accelerometer, M microphone, PS pressure sensor



Comparison of the relationship between F_0 and P_s before and after the operation. The slopes decrease with elevated frequencies both pre- and postoperatively. The slopes of F01 and F02 increase significantly postoperatively (36.43 ± 14.68 pre-operatively, 53.91 ± 30.71 postoperatively, $p = 0.011$ and 26.02 ± 10.71 pre-operatively, 34.85 ± 17.92 postoperatively, $p=0.046$ for F01 and F02, respectively). VHI-10 was also significantly lower postoperatively (16.23 ± 8.00 vs 9.36 ± 6.55 , $p = 0.009$). The preoperative PTP is 0.51 ± 0.16 kPa. The PTP measured 2 weeks after the operation is 0.39 ± 0.11 kPa. The difference between the values before and after surgery is significant ($p = 0.012$).

F_0 fundamental frequency, P_s subglottic pressure, F01 starting frequency, F02 frequency 4 semitones higher than F01, F03 frequency 4 semitones higher than F02. * $p < 0.05$

References:

- Hirano, M. (1975). Phonosurgery, basis and clinical investigation. *Otologica*, 21, 239-240.
- Tanaka, S., & Hirano, M. (1990). Fiberscopic estimation of vocal folds stiffness in vivo using the sucking method. *Arch Otolaryngol Head Surg* 116, 721-724.
- Kazemirad, S., Bakhshaei, H., Mongeau, L., & Kost, K. (2014). Non-invasive in vivo measurement of the shear modulus of human vocal fold tissue. *J Biomech*, 47(5), 1173-1179. doi:10.1016/j.jbiomech.2013.11.034
- Hsiao, T.-Y., Liu, C.-M., Luschei, E. S., & Titze, I. R. (2001). The Effect of Cricothyroid Muscle Action on the Relation Between Subglottal Pressure and Fundamental Frequency in an In Vivo Canine Model. *Journal of Voice*, 15(2), 187-193. doi:10.1016/s0892-1997(01)00020-0
- Alipour, F., & Scherer, R. C. (2015). Time-Dependent Pressure and Flow Behavior of a Self-oscillating Laryngeal Model With Ventricular Folds. *J Voice*, 29(6), 649-659. doi:10.1016/j.jvoice.2014.10.021
- Tseng, W. H., Chang, C. C., Yang, T. L., & Hsiao, T. Y. (2020). Estimating vocal fold stiffness: Using the relationship between subglottic pressure and fundamental frequency of phonation as an analog. *Clin Otolaryngol*, 45(1), 40-46. doi:10.1111/coa.13463

Improving vocal fold paralysis detection from voice recordings using a robust ML model

Juliette Dindart¹, Hélène Massis², Trung Kien Bui¹, Agnès Rouxel^{1,3}, Christophe Trésallet⁴, Claire Pillot-Loiseau², Frédérique Frouin¹

¹ LITO, Inserm, Institut Curie, Université Paris Saclay, Orsay, France

² Laboratoire de Phonétique et Phonologie (LPP), CNRS UMR 7018, Université Sorbonne Nouvelle, Paris, France

³ Nuclear Medicine and ⁴ Endocrine Surgery Departments, Avicenne Hospital, AP-HP, Bobigny, France

Keywords: Vocal Fold Paralysis, machine learning, voice pitch jump

Introduction & rationale:

Following thyroid surgery, about 3 to 5% of patients may experience recurrent nerve damage. This damage can cause vocal fold paralysis (VFP). To ensure that the recurrent nerves are intact, laryngoscopy should be performed before and after surgery, but as this procedure is time-consuming and uncomfortable for patients, it is not always performed during routine care, which can lead to delayed or missed diagnoses.

Vowel phonation recordings could offer an alternative method to detect surgery induced VFP by giving clinicians access to parameters such as jitter, shimmer, and harmonic to noise ratio. Those parameters are known to be altered in patients with recurrent nerve lesions. This simple noninvasive approach could be used as a first sorting of patients.

In a previous work (Dindart et al, 2024), a classification was proposed to separate healthy from VFP voices, based on a machine learning (ML) model. This model was built using a first public database: the Saarbruecken Voice Database (SVD) (Barry et al, 2007) and externally tested using a second public database: the Advanced Voice Function Assessment Database (AVFAD) (Jesus et al, 2017). This model was based on five complementary features: the mean fundamental frequency (f_0), the standard deviation of f_0 (std f_0), the local jitter (lj), the local shimmer (ls), and the harmonic to noise ratio (HNR). To consider inherent variations in the feature values due to various recording acquisitions, the ComBat method was applied to harmonize features on the two databases. In this work, based on the analysis of erroneous classifications, we improve the model by proposing substitution parameters, that are more robust to speech feature extraction algorithms and to noisy data.

Objectives: Propose more robust parameters to better classify between normal and VFP voices

Methods:

For the training/validation step, we selected from the SVD continuously vocalized /a/ vowel of control subjects (428 females and 259 males) and of subjects afflicted with VFP (139 females and 74 males), produced at a normal pitch.

The features used in the model were extracted from the recordings using the Praat library (Python version of Praat), using default parameters. Five features: f_0 , std f_0 , lj, ls, and HNR were selected based on their complementarity. Following a stratified five-fold cross-validation (5CV) procedure, five random forest classifiers were built, using the default values of the scikit-learn library and optimizing the accuracy.

Looking more deeply at classification errors, and computing spectrograms of associated voice records, a jump pitch effect was observed in a few cases (see Figure 1 for an illustrative case, for which the mean fundamental frequency f_0 is 158.9 Hz, while the median fundamental frequency is 174.30 Hz). This modification of pitch was already reported (Vaysse et al, 2022) and induces biased values of f_0 and large values of std f_0 . We thus tested the substitution of f_0 with the median fundamental frequency (f_{050} corresponding to the 50th percentile), and the substitution of std f_0 with the inter percentile range between the 16th and the 84th percentiles (IPR₁₆₋₈₄). These features are less sensitive to extreme values because they are based on percentiles, which makes them more robust.

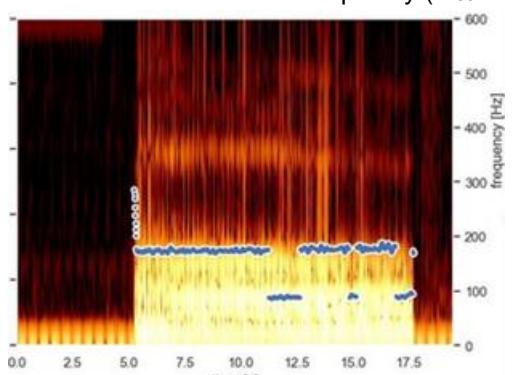


Figure 1: Spectrogram of vocalized /a/ vowel from a healthy subject, with visible pitch jumps

To fully estimate the impact of new features compared to the use of previously selected features, we repeated training/validation procedure defined above using the new features. We tested all the models on an independent external dataset, the AVFAD database, with 396 recordings (363 control subjects and 33 patients with VFP, vocalized /a/ vowel, produced at a normal pitch). To overcome the differences between the features of interest that exist between the two databases (SVD and AVFAD), we also applied the ComBat data harmonization (Behdenna et al. 2023).

In a first experiment, called AVFAD, we used the Praat features provided by the AVFAD project. In a second experiment, considering that the records in the AVFAD database were made of three repetitions of the vowel /a/ phonation, features were computed from each repetition, by splitting the audio file using an automatic method detecting silence intervals. Each model was then tested

on the 1188 (3x396) records and for each subject, the mean value of the three predictor probabilities was computed. From this mean value, the final predicted class was defined. This experiment was called AVFAD₃.

Results:

Table 1 provides the results using both accuracy and balanced accuracy obtained on the validation datasets and on the external database, according to the different experiments. For each experiment, five models, one per fold of the cross-validation, were tested. Thus, eight configurations were compared on the external database: using former or new features, using or not using ComBat harmonization, working with the parameters provided by the AVFAD project or considering the 3 repetitions of the vowel phonation.

Exp.	Metrics	5 CV Validation SVD	5CV Validation SVD ComBat	External Test AVFAD	External Test AVFAD ComBat	External Test AVFAD ₃	External Test AVFAD ₃ ComBat
Former features	Accuracy	0.847±0.012	0.839±0.012	0.611±0.091	0.812±0.023	0.510±0.033	0.765±0.012
	Balanced Accuracy	0.699±0.033	0.690±0.034	0.628±0.039	0.586±0.025	0.689±0.016	0.787±0.012
New features	Accuracy	0.847±0.014	0.836±0.011	0.563±0.034	0.826±0.012	0.575±0.031	0.789±0.015
	Balanced Accuracy	0.702±0.026	0.733±0.015	0.709±0.005	0.726±0.020	0.722±0.017	0.794±0.014

Table 1: Mean values ± standard deviation of accuracy and balanced accuracies obtained for the different experimental configurations (Exp). The two best results for the test database (considering both accuracy and balanced accuracy) are in bold.

On the external test database AVFAD, the ComBat harmonization greatly improves accuracy (from 61.1% to 81.2% for the former features and from 56.3% to 82.6% for the new features), while preserving the balanced accuracy. Moreover, the new features largely improve the balanced accuracy (from 58.6% to 72.6%). The combination of ComBat and the new features yields the best global results (accuracy of 82.6%, balanced accuracy of 72.6%).

Similar trends were observed when considering the 3 repetitions of the vowel phonation. ComBat procedure largely increases the accuracy (from 51% to 76.5% for the former features, from 57.5% to 78.9% for the new features). The combination of ComBat and of the new features further improve the performances of the model, especially in balanced accuracy (accuracy of 78.9%, balanced accuracy of 79.4%).

The separate processing of the three repetitions enables to improve balanced accuracy (increase of 7%) with a decrease of accuracy (minus 4%). It's worth noting that increasing balanced accuracy is crucial to making PFV detection more effective.

Conclusion/Discussion:

The use of median fundamental frequency f₀₅₀ (instead of f₀) and IPR₁₆₋₈₄ (instead of stdf₀) improves the accuracy and balanced accuracy of the classification between controls and VFP. This increase is especially noticeable in balanced accuracy, which is crucial for imbalanced datasets. Largely unbalanced datasets are of course to be expected in a general population undergoing neck surgery.

The use of the ComBat harmonization enables prediction models to be applied to unknown external databases, for which voice recordings are acquired under different, but controlled, conditions.

Finally, the use of repeated acquisitions improved the balanced accuracy, enabling better detection of VFP cases.

References:

- [1] Dindart, J., Massis, H., Bui, T.K., et al (2024 22-26 July). Classification of vocal fold paralysis versus controls based on voice recordings with a Random Forest model: robustness validated using an external database, *13th International Conference on Voice Physiology and Biomechanics*, Erlangen, Germany.
- [2] Barry, W.J., & Pützer, M. (2007) Saarbrücken Voice Database, Institute of Phonetics, Univ. of Saarland, <https://www.stimmdatenbank.coli.uni-saarland.de/>
- [3] Jesus, L.M.T., Belo, I., Machado, J., et al. (2017). The Advanced Voice Function Assessment Databases (AVFAD): Tools for Voice Clinicians and Speech Research. In Fernandes F (Ed.), *Advances in Speech-language Pathology*. Rijeka: InTech. <https://doi.org/10.5772/intechopen.69643>
- [4] Vaysse, R., Astésano, C., & Farinas, J. (2022). Performance analysis of various fundamental frequency estimation algorithms in the context of pathological speech. *Journal of the Acoustical Society of America* 152 (5), 3091-3101. <https://doi.org/10.1121/10.0015143>
- [5] Behdenna, A., Colange, M., Haziza, J., & al. (2023). pyComBat, a Python tool for batch effects correction in high-throughput molecular data using empirical Bayes methods. *BMC Bioinformatics* 24, 459. <https://doi.org/10.1186/s12859-023-05578-5>

This work was partly supported by ANR (ANR-22-CE19-0025).

Use Cases for Optopalatographic Articulatory Measurements: A Portable and Easy to Use Solution

João Menezes, Arne-Lukas Fietkau, Jihyeon Yun, Peter Birkholz

Institute of Acoustics and Speech Communication, Dresden University of Technology, Dresden, Saxony, Germany

Keywords: Optopalatography, speech articulation, use cases

Introduction & rationale: Quantitative measurement of speech articulation is a necessary tool for multiple applications, such as experimental phonetics, basic research on speech production, speech therapy and rehabilitation, and Silent Speech Interfaces (SSIs). Each application has its own necessities, which limits the measurement options possible to be used. For example, as basic research on speech production may privilege precision instead of comfort and portability, studies on experimental phonetics may focus on the portability of the device, especially field studies. Optopalatographic (OPG) measurement devices, and more specifically the OPG devices developed at the Dresden University of Technology (TUD) for the past decade (Birkholz *et al.*, 2023), have their own set of features which make them an interesting alternative to more traditionally applied devices for a broad spectrum of use cases.

Objectives: This abstract aims at presenting published and future use cases for OPG-based articulatory measurements, as well as a functioning device capable of carrying out such measurements.

Methods: OPG articulatory measurements function based on a light emitter and a photodetector: the light emitter emits an optical signal, and the photodetector receives this signals after its reflection by the articulators. We will name this pair of light emitter and photodetector as optode for the remaining of this abstract. Figure 1 shows on its right side the non-personalized and the personalized versions of the OPG2023 device developed at the TUD. They are composed of 13 optodes (15 in the personalized version, with 2 additional for the lip articulation) mounted on a flexible printed circuit board (PCB). The PCB is then connected to a control unit (the black box), which is the interface between the optodes and the recording software, as shown in the left side of Figure 1. After the OPG2023 is worn by a speaker, the optodes face the tongue and the lips and provide measurement values with a sampling rate of 100 Hz based on how distant they are from the articulators.



Figure 1. Left: the simple setup required to record with the OPG2023 (a computer, a control unit and custom software). Right: a non-personalized device on the top, and a personalized device on the bottom.

Results: One of the most frequent use cases of OPG is speech therapy. Already in the decade of 1990, OPG has been used to teach hearing-impaired speakers to produce vowels (Fletcher et al., 1991). More recently it has been applied as a tool to provide visual feedback for motor exercises as a Serious Game (Neuschaefer-Rube et al., 2014) and for post-stroke dysphagia therapy (Wagner et al., 2022). An additional perspective use case for the OPG in speech therapy is being an alternative to electromagnetic articulography (EMA) in the quantification of articulation before and after oral cancer treatment, as in the EMA-study by Tienkamp et al. (2024).

SSIs capable of recognizing command words are another use case for OPG, where speaker dependent accuracies around 98% have been achieved using OPG as sensing modality by Stone & Birkholz (2020), with a Corpus of 30 words, and by Menezes et al. (2024), with a Corpus of 40 words.

Due to the portability of OPG2023, which requires only a small control unit and a regular portable computer to record data, a prospective use case is in experimental phonetics field recordings. Using the non-personalized version of the device, a research group could go on the field and record the articulation of several speakers. The devices can be reused if they are thoroughly cleaned between recordings of different speakers. This can be a viable alternative for recording speech data of under-resourced or endangered languages, which might have its native speakers far from urban-located laboratories.

Multimodal recordings are facilitated by the portability of OPG2023, what widens the horizon of prospective use cases. Recording sessions combining OPG with respiratory speech belts, for example, could provide insight on the coordination of breathing and articulation. Additionally, the combination of OPG with sensing modalities commonly used in SSIs such as surface electromyography (sEMG), ultrasound or radar have the potential to enhance SSI performance and provide direct performance comparison between sensing modalities.

Basic speech production research might also benefit from OPG articulatory measurements. The simultaneous recording of OPG, electroglottography (EGG) and nasal airflow allows to track the influence of voice source and nasal leakage on formant frequencies and bandwidths. While formants are generally considered to be acoustic correlates of tongue position, they are affected by incomplete glottal closure and velopharyngeal opening, resulting in frequency shifts and increased bandwidths. The integration of OPG with other devices enables the isolation of the effects of each articulatory factor on the formants, with the speaker's OPG signals demonstrating a strong correlation with their first two vowel formants. As such, it can improve formant prediction as well as estimation of vocal tract configuration from speech sounds based on articulatory measurements in a non-invasive manner.

Conclusions: OPG articulatory measurements in general, and more specifically the OPG2023 device, display a broad range of use cases from basic speech research to speech technology and rehabilitation. Its main features of portability and ease to use bring OPG to consideration as a viable alternative to other articulatory measurement techniques, given the use case is appropriate.

References:

- Birkholz, P., Stone, S., Wagner, C., Kürbis, S., Wilbrandt, A. & Bosshammer, M. (2023). A review of palatographic measurement devices developed at the TU Dresden from 2011 to 2022. In: *Proc. of the 20th International Congress of Phonetic Sciences (ICPhS 2023)*, 883-887.
- Fletcher, S. G., Dagenais, P. A. & Critz-Crosby, P. (1991). Teaching Vowels to Profoundly Hearing-Impaired Speakers Using Glossometry. *Journal of Speech, Language, and Hearing Research*, 34(4), 943-956. DOI: 10.1044/jshr.3404.94
- Menezes, J., Fietkau, A., Diener, T., Kürbis, S. & Birkholz, P. (2024). A demonstrator for articulation-based command word recognition. In: *Proc. Interspeech 2024*, 2042-2043.
- Neuschaefer-Rube, C., Preuß, S., Eckers, C. & Birkholz, P. (2014). Entwicklung eines OPG-gesteuertes Serious Games als innovatives therapeutisches Hilfsmittel zur Durchführung mundmotorischer Übungen. In: *31. Wissenschaftliche Jahrestagung der DGPP*, Lübeck, Germany.
- Stone, S. & Birkholz, P. (2020). Cross-Speaker Silent-Speech Command Word Recognition Using Electro-Optical Stomatography. In: *ICASSP 2020 – 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Barcelona, Spain, 7849-7853. DOI: 10.1109/ICASSP40776.2020.9053447
- Tienkamp, T. B., Rebernik, T., Halpern, B. M., van Son, R. J. J. H., Wieling, M., Witjes, M. J. H., de Visscher, S. A. H. J. & Abur, D. (2024). Quantifying Articulatory Working Space in Individuals Surgically Treated for Oral Cancer with Electromagnetic Articulography. *Journal of Speech, Language, and Hearing Research*, 67(2), 384-399. DOI: 10.1044/2023_JSLHR-23-00111
- Wagner, C., Stappenbeck, L., Wenzel, H., Steiner, P., Lehnert, B. & Birkholz, P. (2022). Evaluation of a Non-Personalized Optopalatographic Device for Prospective Use in Functional Post-Stroke Dysphagia Therapy. *IEEE Transactions on Biomedical Engineering*, 69(1), 356-365. DOI: 10.1109/TBME.2021.3094415

Clustering Study of Voice Characteristics of Mandarin Speakers in Lifespan

Dai Xiao, Liang Changwei, Lu Yao, Wu Xiyu, Kong Jiangping

Laboratory of Language Sciences and Department of Chinese Language and Literature,
Peking University, Beijing, China

Keywords: Voice characteristics, Mandarin, Lifespan

Abstract

Introduction & rationale: Quantitative analysis in speech science helps reveal the physiological mechanisms of normal voice production and provides support for the diagnosis of voice disorders. With the development of voice analysis technology, Multi-Dimensional Voice Processing (MDVP) has been widely used for the extraction and analysis of acoustic parameters such as fundamental frequency (F0), jitter, and shimmer [1,2,3]. However, research on the voice characteristics of Mandarin speakers across different age and gender groups is still limited, especially in terms of large-scale sampling and multi-parametric analysis.

Objectives: The objective of this research is to describe the voice characteristics of Mandarin speakers, explore the effects of age and gender on multiple voice parameters, and reveal the differences in voice characteristics among different genders and age groups through quantitative analysis.

Methods: This study involved 767 Mandarin speakers, including 402 females and 365 males. Both male and female participants were divided into eight age groups: 3-6 years, 7-10 years, 11-15 years, 16-18 years, 19-30 years, 31-45 years, 46-60 years, and over 60 years, resulting in 16 groups with approximately 50 people in each. Participants pronounced the sustained vowel /a/ at a comfortable pitch and loudness level twice, lasting 2-3 seconds. Each sample was analyzed using the KAY3700 MDVP to extract 34 voice parameters, which were then statistically processed. Pearson correlation analysis, principal component analysis (PCA), K-means clustering, and analysis of variance (ANOVA) were used for statistical analysis [5,6].

Results: Firstly, statistical analysis was performed on the voice parameters of the 16 groups to obtain their respective means, variances, and thresholds. Secondly, cluster analysis was conducted on the 16 groups of samples, with each group's parameters divided into three clusters. This study found that the clustering category characteristics of different age groups and genders were consistent, exhibiting similarity in their parameter structure, as shown in Figure 1. Participants in Cluster 1 had a larger Average Pitch Period (T0), while the Average Fundamental Frequency (F0), Mean Fundamental Frequency (MF0), Highest Fundamental Frequency (Fhi), and Lowest Fundamental Frequency (Flo) were smaller. Cluster 2 exhibited smaller T0 values, with Fo, MF0, Fhi, and Flo significantly greater than the mean of the corresponding age group. The differences in Cluster 3 were more pronounced among different genders and age groups. Finally, ANOVA showed that, except for Number of Unvoiced Segments (NUV), Number of Calculated Segments (SEG), and Pitch Period (PER), the remaining 31 acoustic parameters showed significant differences among different age groups and gender groups ($p < 0.05$).

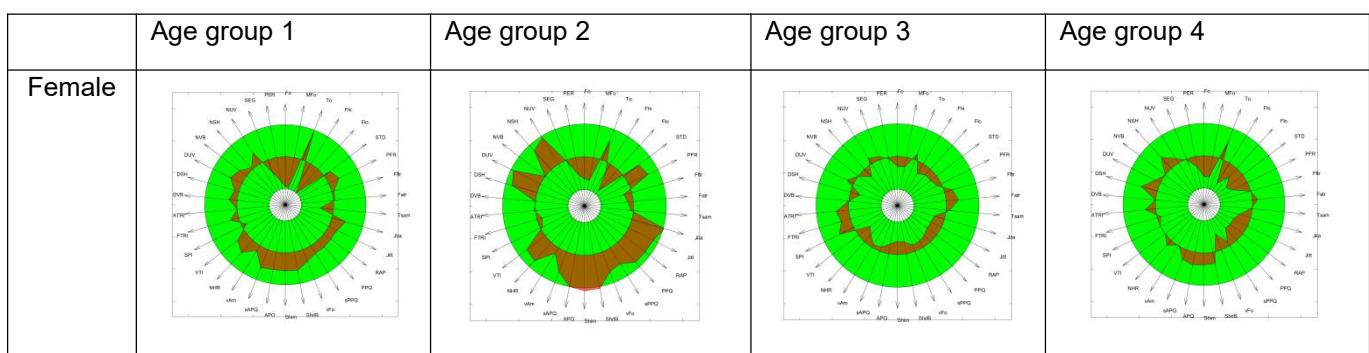




Figure 1: Visualization of Voice Characteristics of Cluster 1 in 16 Groups of Mandarin Speakers

Conclusions: Although the means of Cluster 1, Cluster 2, and Cluster 3 within the 16 groups showed certain similarities relative to the overall means of their respective groups, there were still significant differences in acoustic characteristics among the 16 groups. This study demonstrated the structural differences in acoustic characteristics among 16 groups and revealed significant effects of age and gender on these characteristics. The findings provide empirical evidence for the differences in voice characteristics among Mandarin speakers and lay the foundation for the diagnosis and treatment of voice disorders, emphasizing the need to consider age and gender factors when assessing voice.

Acknowledgments: This research is funded by the Ministry of Education of China. Grant No: 17JJD740001.

References:

- [1] Wu, X. Y., & Kong, J. P. (2020). The lifespan changes of Mandarin phonetics. *Journal of Chinese Phonetics*, 54(2), 17–30.
- [2] Lovato, A., De Colle, W., Giacomelli, L., Piacente, A., Righetto, L., Marioni, G., & de Filippis, C. (2016). Multi-Dimensional Voice Program (MDVP) vs Praat for assessing euphonic subjects: A preliminary study on the gender-discriminating power of acoustic analysis software. *Journal of Voice*, 30(6), 765.e1–765.e5.
- [3] Akil, F., Yollu, U., Ozturk, O., & Yener, M. (2017). Differences of the voice parameters between the population of different hearing thresholds: findings by using the multi-dimensional voice program. *Clinical and experimental otorhinolaryngology*, 10(3), 278-282.
- [4] Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster analysis* (5th ed.). John Wiley & Sons, Ltd.
- [5] Jia, J. P. (2006). *Statistics*. Tsinghua University Press.

ENHANCING GLOTTIS SEGMENTATION IN LARYNGEAL HIGH-SPEED VIDEOENDOSCOPY USING YOLO-BASED ROI DETECTION

Alondra Araya^{1,2}, Emiro Ibarra², Matías Zañartu^{1,2}

¹ Department of Electronic Engineering, Universidad Técnica Federico Santa María , Valparaíso, Chile

² Advanced Center for Electrical and Electronic Engineering, Universidad Técnica Federico Santa María, Valparaíso, Chile

Keywords: Deep Learning, Glottis Segmentation, Laryngeal High-Speed Videoendoscopy

Abstract

Introduction & rationale:

Automatic glottis segmentation in laryngeal high-speed videoendoscopy (HSV) is a fundamental task for biomedical analysis and the assessment of vocal fold-related pathologies. While significant progress has been made using deep learning models based on U-Net (Kruse et al., 2023) and its advanced variants, such as S3AR U-Net (Montalbo, 2024), critical limitations persist, particularly regarding generalization across diverse clinical settings. Although these deep learning models often achieve high performance when trained and tested on a single dataset, their accuracy drops significantly when evaluated on new datasets from other laboratories and different camera setups. This lack of generalization limits their clinical applicability, as HSV recordings exhibit substantial variability in quality, camera settings, and experimental conditions. One key factor that can enhance segmentation performance is the use of regions of interest (ROI). Previous studies have shown that incorporating a well-defined ROI can reduce noise and significantly improve segmentation model performance (Döllinger et al., 2022). However, in many cases, ROI is manually defined or generated using non-fully automatic methods, which introduce biases, require human intervention, and limit scalability. In this study, we explore the application of YOLO (Joseph et al., 2015), renowned for its speed and accuracy in object detection, to automate the generation of ROI in glottis segmentation tasks. This automation aims to reduce manual effort and facilitate the exploration of various ROI application strategies. To assess the generalization capabilities of the proposed method, we conduct cross-dataset validation using two distinct datasets, each encompassing recordings from patients with diverse vocal pathologies and varying camera settings. The results highlight the importance of ROI detection in improving the model's robustness and adaptability across different clinical scenarios.

Objectives:

To study the impact of automating ROI selection using YOLO to improve the accuracy and generalization of deep learning-based glottis segmentation in HSV, evaluated through cross-dataset testing.

Methods:

In this work, we used two datasets: the Benchmark for Automatic Glottis Segmentation (BAGLS) (Gómez et al., 2020) and the Glottal Imaging Dataset for Advanced Segmentation, Analysis, and Facilitative Playbacks Evaluation (GIRAFE) (Andrade-Miranda et al., 2024). The BAGLS dataset comprises 59,250 HSV frames sourced from 640 recordings. In contrast, the GIRAFE dataset includes 32,630 HSV frames from 50 recordings. Both datasets encompass subjects with healthy voices as well as various voice pathologies.

To automatically obtain regions of interest, YOLOv8 was implemented. This network comprises 76 progressive convolutional layers, starting with 16 filters in the initial layer, with the number of filters doubling in each subsequent layer, reaching up to 256 filters. After each convolution, batch normalization and the SiLU activation function are applied to enhance training stability and model performance. The model concludes with a final convolutional layer that generates the segmentation map, also utilizing the SiLU activation function. For training, the model was fine-tuned using 55,750 images from the BAGLS training set over 100 epochs.

For glottis segmentation, two architectures were selected: U-Net, due to its extensive use in state-of-the-art studies (Kruse et al., 2023 and Montalbo, 2024), and S3AR U-Net for its high performance (Montalbo, 2024). In this study, the U-Net consists of four layers in both the encoder and decoder. The encoder starts with 16 filters, doubling the number in each subsequent layer (16, 32, 64, and 128 filters). After each convolution, instance normalization and the ReLU activation function are applied. The decoder follows the same structure, using identical operations and activation functions. At the output, a sigmoid activation function is used for binary segmentation. The S3AR U-Net is an optimized version of U-Net, featuring five layers in both the encoder and decoder. It starts with 4 filters, progressively increasing to 64 filters (4, 8, 16, 32, 64). This model integrates depthwise separable convolutions, attention-gated mechanisms, residual blocks, a feature similarity module, and Squeeze-and-Excitation blocks to enhance segmentation performance while improving computational efficiency. Both models were trained on the BAGLS dataset using 55,750 training images.

To investigate the impact of automated ROI selection on U-Net segmentation performance, two strategies were implemented: 'U-Net with YOLO-based ROI Cropping' and 'U-Net with YOLO-based ROI Filtering.' In the cropping approach, YOLO identified and localized the ROI, which was then extracted by cropping the original image before training U-Net. In the filtering approach, U-Net was trained on the original BAGLS images, and post-segmentation, regions outside the YOLO-detected ROI were discarded.

Results:

Figure 1 illustrates the performance of U-Net, S3AR U-Net, and the two ROI-based segmentation strategies. The left column of Figure 1 shows preliminary results indicating that using ROI detected by YOLO enhances the Intersection over Union (IoU), Precision, and Recall of U-Net segmentation, achieving values very close to those obtained by S3AR U-Net, one of the best-performing networks reported in the literature.

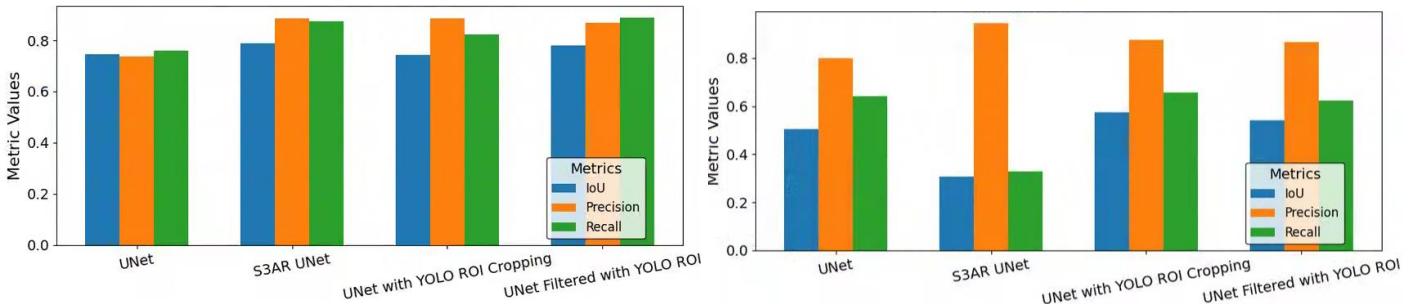


Fig. 1. Performance comparison of models trained on BAGLS. Left: testing on BAGLS; Right: testing on GIRAFE.

However, when evaluating the performance of different networks on the new dataset, a decline in performance is observed, as shown in the right column of Figure 1. This highlights the limited generalization capability of segmentation models under varying video conditions. In this case, S3AR U-Net exhibited the worst performance, while the ROI-based strategies improved all three segmentation metrics compared to other configurations, with the cropping strategy yielding slightly better results.

Even though the drop in all metrics on the GIRAFE dataset can be attributed to variability in camera setups and recording conditions between datasets, another factor is that GIRAFE contains pathologies that differ from those found in BAGLS. To illustrate this, Figure 2 presents the GIRAFE test results grouped by subject condition (healthy, pathological, and unknown). As expected, the best performance was obtained in the healthy group, which has the closest match to BAGLS, in contrast to the other two groups. This analysis also highlights that U-Net using YOLO-based ROI Cropping achieved the best performance.

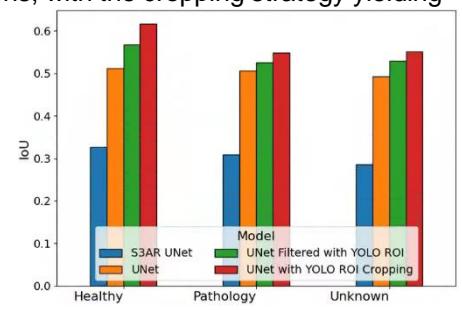


Fig. 2. Test on GIRAFE by subject groups.

Conclusions:

The conducted experiments demonstrate that automatic ROI selection using YOLO enhances glottis segmentation in HSV. However, the drop in metrics during cross-dataset validation across all configurations highlights that while models may achieve outstanding performance on their training dataset, their generalization capability remains limited. Future work will explore advanced domain adaptation techniques, such as transfer learning and adversarial training, to improve model robustness across diverse clinical datasets.

Acknowledgements:

This research was supported by the National Institutes of Health (NIH) National Institute on Deafness and Other Communication Disorders grant P50 DC015446, ANID FONDECYT 1230828, and BASAL AFB240002. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

References:

- Andrade-Miranda, G., Chatzipapas, K., Arias-Londoño, J. D., & Godino-Llorente, J., I. (2024). GIRAFE: Glottal Imaging Dataset for Advanced Segmentation, Analysis, and Facilitative Playbacks Evaluation. *arXiv (Cornell University)*.
- Döllinger, M., Schraut, T., Henrich, L. A., Chhetri, D., Echternach, M., Johnson, A. M., Kunduk, M., Maryn, Y., Patel, R. R., Samlan, R., Semmler, M., & Schützenberger, A. (2022). Re-Training of Convolutional Neural Networks for Glottis Segmentation in Endoscopic High-Speed Videos. *Applied Sciences*, 12(19), 9791.
- Gómez, P., Kist, A. M., Schlegel, P., Berry, D. A., Chhetri, D. K., Dürr, S., Echternach, M., Johnson, A. M., Kniesburges, S., Kunduk, M., Maryn, Y., Schützenberger, A., Verguts, M., & Döllinger, M. (2020). BAGLS, a multihospital Benchmark for Automatic Glottis Segmentation. *Scientific Data*, 7(1).
- Joseph, R., Santosh, D., Ross, G., & Ali, F. (2015). You only look once: Unified, Real-Time Object Detection. *arXiv (Cornell University)*.
- Kruse, E., Dollinger, M., Schutzenberger, A., & Kist, A. M. (2023). GlottisNeTV2: Temporal glottal midline detection using deep convolutional neural networks. *IEEE Journal of Translational Engineering in Health and Medicine*, 11, 137-144.
- Montalbo, F. J. P. (2024). S3AR U-Net: A separable squeezed similarity attention-gated residual U-Net for glottis segmentation. *Biomedical Signal Processing and Control*, 92, 106047.

Visual Perceptual Analysis of Vibratory Patterns and Mucosal Wave Differences Between Vocally Healthy Children and Children with Vocal Fold Nodules

Bernhard Jakubaß¹, Hamide Ghaemi¹, Maggie Lynn Stall², Dimitar D. Deliyski¹, Rita R. Patel²

¹ Department of Communicative Sciences and Disorders, Michigan State University, East Lansing, MI, USA

² Department of Speech, Language and Hearing Sciences, Indiana University, Bloomington, IN, USA

Keywords: *Pediatric, Nodules, HSV*

Abstract

Introduction & rationale: Nodules on the vocal folds in children are a common appearance caused by excessive vocal phonotrauma. Because of the potential of hindering communication, and thus, the development of the child, it is of high importance to understand how nodules affect the phonatory process. (Martins et al., 2012)

In literature, the effect of nodules in children on phonation has been described for the lateral (Patel et al., 2016) and vertical (Patel et al., 2024) motion of the vocal folds. Data describing the vibratory pattern along the longitudinal anterior-posterior (AP) axis (Orlikoff et al., 2012) and the mucosal wave are lacking for children with vocal fold nodules.

Objectives: In the clinic, endoscopic imaging is analyzed through visual perception rather than quantitative measurements. (Döllinger et al., 2012, Poburka et al., 2017) The objective of this study is to conduct a systematic investigation and comparison of the AP vibratory patterns, posterior opening and mucosal wave of vocal fold oscillations in healthy children and children with nodules, as well as between girls and boys.

Methods: High-speed videoendoscopy recordings of the sustained phonations of 40 subjects (10 healthy boys, 10 healthy girls, 10 boys with nodules, and 10 girls with nodules) have been analyzed. For each 4,000-fps video recording, three 62.5-ms sections of stable sustained phonation were selected.

Two experts detected the anterior and posterior commissure points for each section independently. Between the commissures, five equidistant kymogram lines were extracted from the video at 30%, 42.5%, 55%, 67.5%, and 80% of the A-P distance. The AP vibratory patterns (Orlikoff et al., 2012) and the mucosal wave (Patel et al., 2018) were rated. This analysis was done by the two expert raters independently, inter-rater reliability was evaluated and a third rater mediated ad-hoc consensus.

Results: The presentation discusses the AP vibratory patterns, posterior glottal gap and mucosal wave ratings, as well as the statistical analysis results, comparing the findings between the four groups. The results are discussed in comparison to other vibratory patterns and mucosal wave descriptions in literature.

Conclusions: This study complements previous research (Patel et al., 2016, 2024) to fully understand and describe the changes within the vocal fold vibratory movement due to nodules in children in all dimensions.

References:

- Döllinger, M., Dubrovskiy, D., & Patel, R. R. (2012). Spatiotemporal analysis of vocal fold vibrations between children and adults. *The Laryngoscope*, 122(11), 2511–2518. <https://doi.org/10.1002/lary.23568>
- Martins, R. H., Hidalgo Ribeiro, C. B., Fernandes de Mello, B. M., Branco, A., & Tavares, E. L. (2012). Dysphonia in children. *Journal of voice : official journal of the Voice Foundation*, 26(5), 674.e17–674.e6.74E20. <https://doi.org/10.1016/j.jvoice.2012.03.004>
- Orlikoff, R. F., Golla, M. E., & Deliyski, D. D. (2012). Analysis of longitudinal phase differences in vocal-fold vibration using synchronous high-speed videoendoscopy and electroglottography. *Journal of voice : official journal of the Voice Foundation*, 26(6), 816.e13–816.e8.16E20. <https://doi.org/10.1016/j.jvoice.2012.04.009>
- Patel, R. R., Unnikrishnan, H., & Donohue, K. D. (2016). Effects of Vocal Fold Nodules on Glottal Cycle Measurements Derived from High-Speed Videoendoscopy in Children. *PloS one*, 11(4), e0154586. <https://doi.org/10.1371/journal.pone.0154586>
- Patel, R. R., Awan, S. N., Barkmeier-Kraemer, J., Courey, M., Deliyski, D., Eadie, T., Paul, D., Švec, J. G., & Hillman, R. (2018). Recommended Protocols for Instrumental Assessment of Voice: American Speech-Language-Hearing Association Expert Panel to Develop a Protocol for Instrumental Assessment of Vocal Function. *American journal of speech-language pathology*, 27(3), 887–905. https://doi.org/10.1044/2018_AJSLP-17-0009
- Patel, R. R., Döllinger, M., & Semmler, M. (2024). 3D reconstruction of vocal fold dynamics with laser high-speed videoendoscopy in children. *Laryngoscope investigative otolaryngology*, 9(5), e70024. <https://doi.org/10.1002/lio2.70024>

Poburka, B. J., Patel, R. R., & Bless, D. M. (2017). Voice-Vibratory Assessment With Laryngeal Imaging (VALI) Form: Reliability of Rating Stroboscopy and High-speed Videoendoscopy. *Journal of voice : official journal of the Voice Foundation*, 31(4), 513.e1–513.e14. <https://doi.org/10.1016/j.jvoice.2016.12.003>

Feedback and Feedforward Auditory-Motor Processes for Voice and Articulation in Hearing Loss

Andrea Alves Maia^{1,2}, Luis M. T. Jesus²

¹ Department of Speech Therapy – Health Sciences Center – Federal University of Espírito Santo, Brazil

² School of Health Sciences (ESSUA), Institute of Electronics and Informatics Engineering of Aveiro (IEETA), Intelligent Systems Associate Laboratory (LASI), University of Aveiro, Portugal

Keywords: Auditory-motor integration; Suboptimal auditory feedback; Cochlear implant

Introduction

Auditory-motor processing involves the combination of feedback and feedforward control systems of speech (Guenther, 2006). In the case where speech errors are sustained over time, the feedforward control system produces adaptive responses, responsible for updating the motor plan based on corrective commands from the feedback system (Abur et al., 2021).

In cochlear implant (CI) users, although auditory feedback is partially restored, the information transfer remains suboptimal for perceiving fine spectrotemporal details (Zeng et al., 2008). There is compelling evidence that auditory feedback via CIs allows the fine-tuning of motor planning and execution in speech production, however studies present contradictory findings and do not exhibit a consistent pattern for voice and articulation. Ashjaei et al. (2024) reviewed studies of vocal control and speech production in CI recipients, suggesting that precise control of vocal pitch is highly dependent on auditory feedback from the CI. Despite some mixed findings, the overall trend was that the vowel space changed to improve phonemic contrasts after implantation, towards typical formant frequency values.

Objectives

To evaluate adaptive feedforward responses in the control of the auditory-motor process of voice and articulation in adults with altered feedback due to hearing loss (HL) before (PRE) and after (POS) cochlear implantation.

Method

Speech production of 15 participants (5 women and 10 men) with normal hearing (NH) was analysed, and 10 participants (5 women and 5 men) with severe-profound sensorineural post lingual HL, were evaluated before and after cochlear implantation. The participants' first language was Brazilian Portuguese, the NH group was aged 19-64 years ($M = 41.86$; $SD = 5.75$) without voice disorders according to the Voice Symptom Scale (Moreti et al., 2011), the participants with HL were aged 19-74 years ($M = 47.3$; $SD = 16.08$) PRE cochlear implantation, and aged 24-78 years ($M = 49.7$; $SD = 16.30$) POS cochlear implantation (18.5 ± 17.35 months of CI use). Ethical approval was obtained from an independent Research Ethics Committee in Brazil (6.759.381).

Data was collected in a quiet room, using the Computerized Speech Lab (CSL™) Model 4500, with a Shure PG48 dynamic microphone positioned at a 45° angle from the participants' mouth and 5 cm away from their lips. The recordings were sampled at 44100 Hz with 16-bit per sample. Individuals were instructed to produce a sustained /a/ and read the Consensus Auditory-Perceptual Evaluation of Voice – CAPE-V (Behlau et al., 2022) sentences, in usual and comfortable voice. The acoustic signal was imported into Praat 6.4.08, a TexGrid object with one interval tier was created, and all the productions of oral vowel /a/ were manually annotated using perceptual and acoustic evidence (number of tokens available for analysis: NH = 237; PRE = 172; POS = 175).

The adaptive feedforward responses were evaluated using a method based on previous studies (Lester-Smith et al., 2020; Tourville et al., 2008), with the adaptation of baseline control measures being extracted from the NH group, and the responses to two conditions of sustained auditory feedback perturbations inferred from speech production of participants with HL, PRE and POS cochlear implantation. Mean fundamental (f_0) and first formant (F_1) frequencies were estimated from a 40-120 ms window, after the onset, for all productions of vowel /a/. The f_0 values in Hz were converted to Cents relative to the f_0 mean of the NH group, as a correlate of laryngeal articulation and voiced quality control. The F_1 values were extracted with linear predictive coding and the percent change relative to the F_1 mean of the NH group was calculated, as correlates of oral articulation adjustments during speech production.

R version 4.4.2 running in RStudio Version 2024.12.0+467 was used for statistical analysis, mixed-effects modelling and data visualisation. The models' predictions and shading spanning the 95% confidence interval were drawn using the sjPlot 2.8.16 package. A mixed effects regression model was developed using the `lmer` function from the lme4 version 1.1-35.3 package, with the f_0 and F_1 as outcome variables, considering `group_id` (PRE-POS) and `phase_id` (PRE-NH or POS-NH) as fixed effects, `speaker_id` and `sex_id` as random effects. Results from likelihood ratio tests of the models with the `group_id` and `phase_id` effects against the models without the `group_id` and `phase_id` effects were also analysed.

Results

The average adaptive f_0 response was 43.3 Cents ($SD = 510.5$ Cents) PRE cochlear implantation, and -19.3 Cents ($SD = 553.2$ Cents) POS cochlear implantation. The adaptive PRE F_1 response was 13.6 % ($SD = 14.5$ %), and 4.1 % ($SD = 25.4$ %) POS cochlear implantation. The f_0 and F_1 raw values are shown in figure 1. Two mixed effects regression models with the lme4 syntax $f_0 \sim \text{group_id} + (1|\text{speaker_id}) + (1|\text{sex_id})$ and $F_1 \sim \text{group_id} + (1|\text{speaker_id}) + (1|\text{sex_id})$ predicted the values also shown in figure 1. Likelihood ratio tests of the models with the `group_id` effect against the models without the `group_id` effect revealed a significant difference between

models, i.e., there was a significant difference between f_0 and F_1 values before and after cochlear implantation: $f_0 - \chi^2(1) = 5.345$, $p = 0.021$; $F_1 - \chi^2(1) = 0.005$, $p < 0.001$.

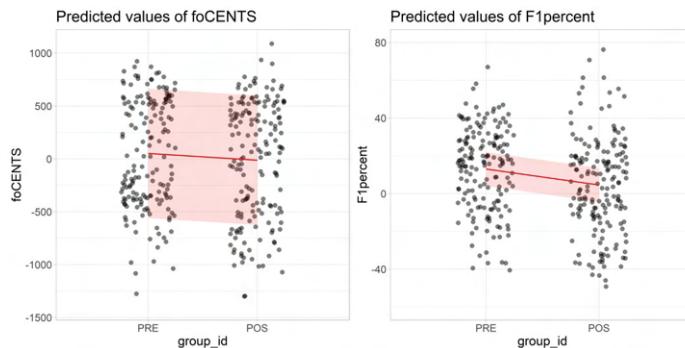


Figure 1. f_0 and F_1 raw values (grey dots) for the two groups (PRE and POS); f_0 and F_1 predicted values (mixed effects regression model regression lines and shading spanning the 95% confidence intervals) are also shown.

Two additional mixed effects regression models with the lme4 syntax $f_0 \sim \text{phase_id} + (1|\text{speaker_id}) + (1|\text{sex_id})$ and $F_1 \sim \text{phase_id} + (1|\text{speaker_id}) + (1|\text{sex_id})$ were developed. Likelihood ratio tests of the models with the `phase_id` effect against the models without the `phase_id` effect only revealed a significant difference between the F_1 PRE-NH models, i.e., there was only a significant difference between PRE cochlear implantation and NH F_1 values: f_0 (PRE-NH) – $\chi^2(1) = 0.025$, $p = 0.874$; F_1 (PRE-NH) – $\chi^2(1) = 5.180$, $p = 0.023$; f_0 (POS-NH) – $\chi^2(1) = 0.530$, $p = 0.467$; F_1 (POS-NH) – $\chi^2(1) = 0.258$, $p = 0.611$.

Conclusions

Adaptive responses of the feedforward system for laryngeal auditory-motor control, in conditions of altered auditory feedback due to severe-profound postlingual HL in adulthood or due the use of CI, promoted adjustments like those of NH (the f_0 PRE-NH and POS-NH models were not significantly different), revealing adaptations of laryngeal motor control through somatosensory feedback. In adults with postlingual HL, the CI provides updates of the articulation motor plan, with adaptive responses and corrective commands from the suboptimal feedback system provided by the CI, allowing adjustments towards typical F_1 values.

Adaptive modifications to auditory feedback yield robust response in speakers with typical speech, in which they compensate by opposing the direction of the perturbation for both vocal and articulatory features of speech (Burnett et al., 1998; Cai et al., 2011; Tourville et al., 2008; Villacorta et al., 2007). Therefore, the lowering of f_0 and F_1 values we observed in adults POS cochlear implantation, suggests a change in their auditory feedback condition with less perturbation for high frequencies, when compared to the PRE cochlear implantation condition.

The feedforward control system is responsible for updating the motor plan based on the corrective commands from the feedback system. The analysis of voice and speech outputs can help understand and define CI programming adjustments for better auditory motor control.

References

- Abur et al. (2021). Feedback and Feedforward Auditory-Motor Processes for Voice and Articulation in Parkinson's Disease. *J SPEECH LANG HEAR R*, 64(12), 4682-4694.
- Ashjaei et al. (2024). Vocal control and speech production in cochlear implant listeners: A review within auditory-motor processing framework. *HEARING RES*, 453, 109132.
- Behlau et al. (2022). Validation of the Brazilian Portuguese CAPE-V Instrument. *J VOICE*, 36(4), 586.e15-586.e20.
- Burnett et al. (1998). Voice F0 responses to manipulations in pitch feedback. *J ACOUST SOC AM*, 103(6), 3153-3161.
- Cai et al. (2011). Focal Manipulations of Formant Trajectories Reveal a Role of Auditory Feedback in the Online Control of Both Within-Syllable and Between-Syllable Speech Timing. *J NEUROSCI*, 31(45), 16483-16490.
- Guenther (2006). Cortical interactions underlying the production of speech sounds. *J COMMUN DISORD*, 39(5), 350-365.
- Lester-Smith et al. (2020). The Relation of Articulatory and Vocal Auditory–Motor Control in Typical Speakers. *J SPEECH LANG HEAR R*, 63(11), 3628-3642.
- Moreti et al. (2011). Equivalência cultural da versão Brasileira da Voice Symptom Scale: VoiSS. *Jornal da Sociedade Brasileira de Fonoaudiologia*, 23(4), 398-400.
- Tourville et al. (2008). Neural mechanisms underlying auditory feedback control of speech. *NEUROIMAGE* 39(3), 1429-1443.
- Villacorta et al. (2007). Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception. *J ACOUST SOC AM*, 122(4), 2306-2319.
- Zeng et al. (2008). Cochlear Implants: System Design, Integration, and Evaluation. *IEEE REV BIOMED ENG*, 1, 115-142.

A Dosimetry Comparative Study in Dear Evan Hansen

Ana Zuim¹, Celia Stewart², Ingo Titze³, Aaron Johnson⁴

¹ Music and Performing Arts Professions/Steinhardt School, New York University, NY, NY, USA

² Communicative Sciences and Disorders/Steinhardt School, New York University, NY, NY, USA

³ Utah Center for Vocology, University of Utah, and National Center for Voice and Speech, Salt Lake City, Utah, USA

⁴ Department of Otolaryngology - Head and Neck Surgery, New York University Grossman School of Medicine, NY, NY, USA

Keywords: dosimetry, vocal dose, vocal load

Introduction and Rationale:

Contemporary musical theater combines songs and dialogue to push boundaries, creating dramatic plots that address current issues. These scores frequently require singers to use concentrated vocal energy below the passaggio. A singer's technique and vocal choices, such as the vocal intensity and registration, influence the vocal dose each performer accumulates during their performances. This integration of music and movement increases the emotional and physiological demands placed on performers. To maintain consistent performance quality over time, ensure vocal longevity, and facilitate ease of singing, singers must understand their "vocal dose."

Performers view modern musicals such as "Dear Evan Hansen" as incredibly challenging due to the demands on singers to uphold their vocal health while balancing their social lives. On Broadway, actors typically perform the same role about eight times a week, which can be very demanding both physically and vocally. In interviews, Ben Platt, the original Evan Hansen, has shared that he greatly limited his voice usage outside of performances to protect his vocal health and manage his busy schedule. Leading up to the Tony Awards in 2017, he was put on vocal rest to enhance his vocal performance, quickly resuming his normal speaking and singing activities after just a few days.

Despite the common anecdotal evidence that roles like Evan Hansen and other main characters such as Jared Kleinman, Connor Murphy, Zoe Murphy, and Alana Beck in "Dear Evan Hansen" are vocally demanding, there is a lack of quantitative information on the actual vocal dose of these roles. Furthermore, the influence of technical vocal choices of individual performers on vocal dose is unknown. Understanding the actual vocal load of these demanding roles and the variability between individual performers could guide recommendations for vocal health and wellness. Thus, our study aimed to document the vocal doses of performers in "Dear Evan Hansen," whether on Broadway or within touring productions. Additionally, our secondary objective was to explore variability of time, cycle, distance doses, and sound pressure levels within roles by comparing dosimetry data from two performers who have portrayed the same role.

Objectives:

The primary objective of this study was to quantify the vocal demands placed on the main characters (Evan Hansen, Jared Kleinman, Alana Beck and Zoe Murphy) in the musical "Dear Evan Hansen" using voice dosimetry. Additionally, we aimed to compare the time, cycle, distance doses, and sound pressure level dosimetry data between performers for each role.

Methods:

Two male-identifying and two female-identifying singers participated in our study; each singer performed multiple roles in a Broadway or touring production of Dear Evan Hansen. Recordings were made in an IAC soundproof booth using a KayPENTAX APM 3300 dosimeter to measure vocal fold vibrations. The singers recorded the musical's songs, followed by separate dialogue recordings to distinguish the vocal demands of singing and speaking.

Results:

Dosimetric analysis of four performers from Dear Evan Hansen identified the distance dose, cycle dose, time dose, and SPL across all compared roles (Evan Hansen, Jared Kleinman, Zoe Murphy, and Alana Beck) across participants. Singers accumulated differing levels of vocal dose when performing the same role, including a larger overall distance dose for those with higher mean SPL.

Singer	Role	Mode	Time Dose (%)	in Meters	Cycle Dose	Pressure
#1M	Evan Hansen	Singing	14 Min, 19 Sec	1384 M	221,889 cy	94 dB
#2M	Evan Hansen	Singing	15 Min, 3 Sec	1410 M	229,227 cy	95 dB
#1M	Evan Hansen	Dialogue	14 Min, 20 Sec	915 M	200,824 cy	85 dB
#2M	Evan Hansen	Dialogue	12 Min, 32 Sec	942 M	166,439 cy	90 dB
#1M	Jared Kleinman	Singing	5 Min, 52 Sec	547 M	108,631 cy	94 dB
#2M	Jared Kleinman	Singing	4 Min, 29 Sec	660 M	79,537 cy	105 dB
#1M	Jared Kleinman	Dialogue	4 Min, 8 Sec	213 M	54,778 cy	84 dB
#2M	Jared Kleinman	Dialogue	4 Min, 1 Sec	266 M	52,969 cy	88 dB
#1F	Zoe Murphy	Singing	7 Min, 42 Sec	656 M	169,120 cy	90 dB
#2F	Zoe Murphy	Singing	7 Min, 56 Sec	596 M	173,962 cy	86 dB
#1F	Zoe Murphy	Dialogue	2 Min, 50 Sec	153 M	44,920 cy	81 dB
#2F	Zoe Murphy	Dialogue	2 Min, 42 Sec	110 M	40,039 cy	74 dB
#1F	Alana Beck	Singing	5 Min, 39 Sec	755 M	134,547 cy	99 dB
#2F	Alana Beck	Singing	6 Min, 37 Sec	523 M	141,435 cy	90 dB
#1F	Alana Beck	Dialogue	4 Min, 6 Sec	341 M	80,129 cy	89 dB
#2F	Alana Beck	Dialogue	3 Min, 34 Sec	230 M	75,214 cy	80 dB

Conclusions/Discussion:

The greater the distance dose, the higher the vocal demand for performers, as it accounts for both amplitude and frequency, making it the most accurate measure of vocal dose. Performers with higher mean fundamental frequencies or elevated SPL accumulate greater vocal fold collision forces. While actors use their natural fundamental frequency for dialogue, variation in this frequency leads to differences in distance dose across performers. All performers exhibited shorter distance doses during dialogue compared to singing.

When either the cycle dose or SPL increases, the total distance dose also rises. This relationship was demonstrated by the singers portraying Jared Kleinman. Singer #M2 performed at 105 dB for approximately 80,000 cycles, while singer #M1 sang at 84 dB for 109,000 cycles. This resulted in a 17% greater distance dose for singer #M2 due to the louder sound.

Female singers showed similar trends. Singer #F2 performed as Zoe Murphy at 86 dB with 174,000 cycles, while singer #F1 sang at 90 dB with 169,000 cycles, resulting in a 9% difference in total distance dose. These observations suggest that the loudness of an individual's singing has a more significant impact on the total distance dose than the number of cycles produced. This insight may aid researchers in establishing thresholds for safe vocal usage during contemporary musical theater productions.

References:

- Bourne, T., & Kenny, D. (2016). Vocal qualities in music theater voice: Perceptions of expert pedagogues. *Journal of Voice: Official Journal of the Voice Foundation*, 30(1). <https://doi.org/10.1016/j.jvoice.2015.03.008>
- Manfredi, C., & Dejonckere, P. H. (2014). Voice dosimetry and monitoring, with emphasis on professional voice diseases: Critical review and framework for future research. *Logopedics Phoniatrics Vocology*, 1–17. <https://doi.org/10.3109/14015439.2014.970228>
- Misono, S., Banks, K., Gaillard, P., Goding, G. S., Jr, & Yueh, B. (2015). The clinical utility of vocal dosimetry for assessing voice rest. *The Laryngoscope*, 125(1), 171–176. <https://doi.org/10.1002/lary.24887>
- Morawska, J., Niebudek-Bogusz, E., & Pietruszewska, W. (2022). Considerations and demands in the voice care of contemporary commercial singers in occupational health and safety aspects. *Medycyna Pracy*, 73(1), 33–41. <https://doi.org/10.13075/mp.5893.01201>
- Santos, S. S., Montagner, T., Bastilha, G. R., Frigo, L. F., & Cielo, C. A. (2019). Singing style, vocal habits, and general health of professional singers. *International Archives of Otorhinolaryngology*, 23(4), e445–e450. <https://doi.org/10.1055/s-0039-1693140>
- Titze, I.R., Svec, J.G., & Popolo, P.S. (2003). Vocal dose measures: Quantifying accumulated vibration exposure in vocal fold tissues. *J Speech Lang Hear Res*. 46(4):919-932. doi:10.1044/1092-4388(2003/072)
- Zuim, A.F., Stewart, C.F., & Titze, I.R. (2021). Vocal dose and vocal demands in contemporary musical theatre. *Journal of Voice: Official Journal of the Voice Foundation*, Oct 4:S0892-1997(21)00278-2. doi:10.1016/j.jvoice.2021.08.006. Epub ahead of print. PMID: 34620516.

Voice cepstral peak prominence metrics in Italian by age, sex, and pre-processing method

Irene de Nijs¹, Thomas Tienkamp^{1,2,3}, Martijn Wieling^{1,3}, Defne Abur^{1,3}

¹ Center for Language and Cognition Groningen, University of Groningen, Groningen, the Netherlands

² Department of Oral and Maxillofacial Surgery, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands

³ Research School of Behavioral and Cognitive Neurosciences, University of Groningen, Groningen, the Netherlands

Keywords: Cepstral peak prominence; voice acoustics

Abstract

Introduction & rationale: Smoothed cepstral peak prominence (CPPS) is an acoustic measure that has been associated with perceptual voice quality. Unlike traditional acoustic time-based voice measures (e.g., jitter/shimmer), it is frequency-based and can be extracted from connected speech, which makes it a promising marker for distinguishing typical speakers from individuals with voice disorders (Murton et al., 2020). CPPS values are often extracted with voicing activity detection (VAD), which tends to yield higher values than those extracted without VAD (Buckley et al., 2023). An advantage of using VAD is that it minimizes variability between speakers by removing pauses and unvoiced segments (Kitayama et al., 2020). However, CPPS values computed with VAD may not accurately reflect dysphonic speech since aphonic segments are removed from the analysis (Awan et al., 2009). Solely removing pauses could solve these issues, since that approach was found to decrease intertext variability and enhance diagnostic accuracy in Japanese speakers with dysphonia (Kitayama et al., 2020). Age has not been found to have a significant effect on CPPS values (Buckley et al., 2023; Yanushevskaya et al., 2024). Regarding sex, males tend to have higher CPPS values than females during sustained vowel phonation (Awan et al., 2012; Lee et al., 2018; Yanushevskaya et al., 2024). During passage reading, however, previous work has found the reverse effect –where females had higher CPPS values compared to males– (Lee et al., 2018) or no significant effect of sex (Taylor et al., 2020). Moving toward the use of CPPS in clinical settings, it is therefore important to determine the separate and combined effects of pre-processing methods, age, and sex.

Objectives: The current study aimed to examine the effect of pre-processing methods, age, and sex on CPPS values in typical Italian speakers. The study expands on Kitayama et al.'s (2020) work by assessing a different language and by including a larger sample of typical speakers with a broader age range.

Methods: Typical Italian speakers ($n=120$) aged 17 to 74 years old were recorded in a sound-attenuated room while they read an Italian translation of "The Rainbow passage" (Fairbanks, 1960). They were recorded at a sampling frequency of 44.1 KHz using a Shure MX153 omnidirectional mic placed 7cm from the mouth at a 45-degree angle. Analyses were completed using four separate pre-processing methods. For the first analyses ("pause removal" or PR) pauses were removed from the recordings using a custom-written script in MATLAB (version 2023b; The MathWorks Inc., 2023). For the second analyses ("voicing activity detection" or VAD), CPPS values were extracted with and without VAD from recordings with and without pauses removed using a batch script (Murray et al., 2022) in Praat (version 6.4.18; Boersma & Weenink 2025). Thus, to assess the effects of pre-processing, the two approaches were compared individually and combined, resulting in four methods. A linear mixed effects analysis was conducted using CPPS as the response variable and methodology ("with/without VAD" or $[+]/[-]$ VAD and "with/without PR" or $[+]/[-]$ PR), age (17-74 years), and sex (male or female) as fixed effects. Subject was coded as a random-effect factor and the alpha value was set at 0.05. The CPPS values were log-transformed, as the residuals were not normally distributed and the homoscedasticity assumption was not met. Post-hoc pairwise comparisons were performed and were corrected for multiple comparisons using the false discovery rate (FDR) procedure.

Results: Figure 1 shows the distribution of CPPS values per sex across the four pre-processing methods, which varied in whether VAD was used ($[+]/[-]$ VAD) and whether pauses were removed ($[+]/[-]$ PR). The final model revealed significant effects of pre-processing method and sex. The $[+]$ VAD & $[+]$ PR method yielded significantly lower CPPS values than the $[+]$ VAD & $[+]$ PR ($\beta = -.28$, $p < .0001$), the $[+]$ VAD & $[+]$ PR ($\beta = -.31$, $p < .0001$) and the $[+]$ VAD & $[+]$ PR methods ($\beta = -.33$, $p < .0001$). There were also significantly lower CPPS values for the $[+]$ VAD & $[+]$ PR method compared to the $[+]$ VAD & $[+]$ PR ($\beta = -.03$, $p < .0001$) and the $[+]$ VAD & $[+]$ PR methods ($\beta = -.05$, $p < .0001$). Finally, CPPS values were significantly lower for the $[+]$ VAD & $[+]$ PR method than for the $[+]$ VAD & $[+]$ PR method ($\beta = -.02$, $p < .0001$). As for sex, males had significantly lower CPPS values than males ($\beta = -.09$, $p < .001$). There was no significant effect of age ($p = .24$).

Conclusions: Significantly higher CPPS values were observed for recordings with pauses removed than for unprocessed recordings, which is in line with previous findings (Kitayama et al., 2020). Diverging from Kitayama et al.'s (2020) work, the current study included the $[+]$ VAD & $[+]$ PR condition, which was found to yield significantly different CPPS values than the three other pre-processing methods.

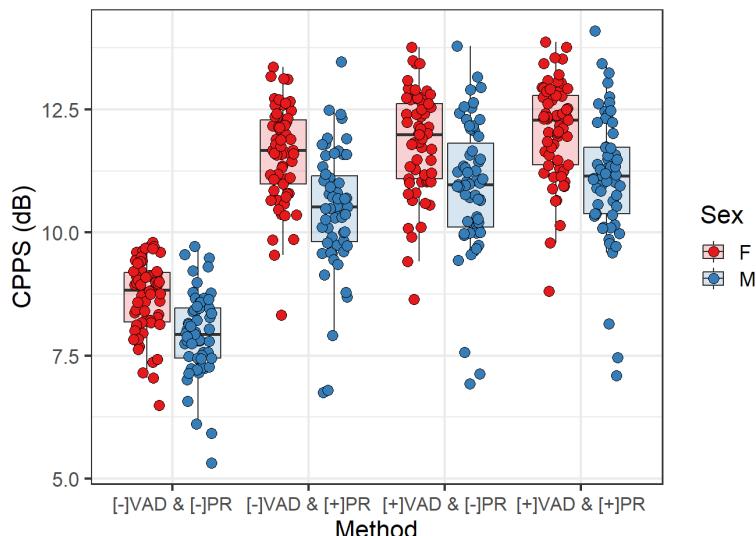


Figure 1. Distribution of smoothed cepstral peak prominence (CPPS) values per sex (F = female, M = male) across four pre-processing methods (VAD = voicing activity detection, PR = pause removal).

Furthermore, the results indicate that the effect of pre-processing method holds up in a larger sample of Italian speakers with a broader age range. The absence of a significant effect of age was consistent with previous findings (Buckley et al., 2023; Yanushevskaya et al., 2024). Finally, males having higher CPPS values than females is in line with a previous study that also analyzed passage readings (Lee et al., 2018), but not with studies that analyzed sustained vowels (Awan et al., 2012; Lee et al., 2018; Yanushevskaya et al., 2024). This could mean that sex differences in CPPS may depend on the stimuli used. The significant effects of methodology and sex on CPPS values in typical Italian speakers indicate that these factors should be considered when using CPPS as a diagnostic marker for voice disorders. Future studies could examine the effect of methodology and sex in individuals with voice disorders.

References

- Awan, S. N., Giovinco, A., & Owens, J. (2012). Effects of vocal intensity and vowel type on cepstral analysis of voice. *Journal of Voice*, 26(5), 670.e15-670.e20. <https://doi.org/10.1016/j.jvoice.2011.12.001>
- Awan, S. N., Roy, N., & Dromey, C. (2009). Estimating dysphonia severity in continuous speech: Application of a multi-parameter spectral/cepstral model. *Clinical Linguistics & Phonetics*, 23(11), 825–841. <https://doi.org/10.3109/02699200903242988>
- Boersma, P. & Weenink, D. (2025). Praat: doing phonetics by computer [Computer program]. Version 6.4.18, retrieved from <http://www.praat.org/>
- Buckley, D. P., Abur, D., & Stepp, C. E. (2023). Normative values of Cepstral peak prominence measures in typical speakers by sex, speech stimuli, and software type across the life span. *American Journal of Speech-Language Pathology*, 32(4), 1565–1577. https://doi.org/10.1044/2023_ajslp-22-00264
- Fairbanks, G. (1960). *Voice and articulation drillbook* Harper & Row.
- Kitayama, I., Hosokawa, K., Iwahashi, T., Iwahashi, M., Iwaki, S., Kato, C., Yoshida, M., Umatani, M., Matsushiro, N., Ogawa, M., & Inohara, H. (2018b). Intertext variability of smoothed cepstral peak prominence, methods to control it, and its diagnostic properties. *Journal of Voice*, 34(3), 305–319. <https://doi.org/10.1016/j.jvoice.2018.09.021>
- Lee, S. J., Pyo, H. Y., & Choi, H. (2018). Normative data of cepstral and spectral measures in Korean adults using vowel phonation and passage reading tasks. *Eon'eo Cheong'gag Jang'ae Yeon'gu/Communication Sciences & Disorders*, 23(1), 208–217. <https://doi.org/10.12963/csd.18474>
- The MathWorks Inc. (2023). MATLAB [Computer program]. Version R2023b, retrieved from https://www.mathworks.com/?s_tid=gn_logo
- Murton, O., Hillman, R., & Mehta, D. (2020b). Cepstral Peak prominence values for Clinical voice evaluation. *American Journal of Speech-Language Pathology*, 29(3), 1596–1607. https://doi.org/10.1044/2020_ajslp-20-00001
- Murray, E. S. H., Chao, A., & Colletti, L. (2022). A practical guide to calculating Cepstral peak prominence in Praat. *Journal of Voice*. <https://doi.org/10.1016/j.jvoice.2022.09.002>
- Taylor, S., Dromey, C., Nissen, S. L., Tanner, K., Eggett, D., & Corbin-Lewis, K. (2020). Age-Related Changes in Speech and Voice: Spectral and cepstral measures. *Journal of Speech Language and Hearing Research*, 63(3), 647–660. https://doi.org/10.1044/2019_jslhr-19-00028
- Yanushevskaya, I., O'Donnell, N., & O'Regan, S. (2024). Cepstral Peak Prominence in normophonic Irish-English speaking adults: The effect of gender, age, speech task segmental composition, recording conditions and CPP extraction method. *Clinical Linguistics & Phonetics*, 1–30. <https://doi.org/10.1080/02699206.2024.2428177>

Exploring the use of amniotic materials to prevent vocal fold scar after injury

Jenny L Pierce¹, Brendan Olson², Ray M Merrill³, Anika Isom⁴, Kim Lanaghen¹, Omar Gharbia¹, Sidney Kaufmann², Marshall E Smith¹, Michael B Christensen^{1,2}

¹Department of Otolaryngology – Head & Neck Surgery, University of Utah School of Medicine, SLC, UT, USA

²Department of Surgery, Division of Urology, University of Utah School of Medicine, SLC, UT, USA

³Department of Communication Sciences & Disorders, Brigham Young University, Provo, UT, USA

⁴Department of Biology, University of Utah, SLC, UT, USA

Keywords: Scar; Amniotic; Inflammation

Abstract

Introduction & rationale: Vocal fold (VF) scarring results in low quality of life and is notoriously difficult to treat. Hyaluronic acid-based (HA) injectables are often used to treat vocal fold injury, but the effects are temporary. Amniotic fluid (AF) and amniotic membrane (AM) have demonstrated healing potential in a variety of organs, including the airway. Therefore, we investigated the use of amniotic materials in a VF scar model.

Objectives: To determine the ability of various amniotic materials to prevent or reduce vocal fold scarring following injury.

Methods: For Phase 1 of this study, six groups of 10 (N=60) male New Zealand white rabbits had a punch biopsy of one VF, with the contralateral VF serving as an internal uninjured control. Treatment (HA, AF, or sham/saline) was injected immediately into the wound bed following injury. Each treatment group was analyzed at four and 10 weeks post-injury (n=30 each). Rheology analysis included viscous and elastic moduli. RT-qPCR analysis included TNF- α , IL-1 β , and IL6 pro-inflammatory cytokines.

For Phase 2, seven treatment conditions were investigated: non-surgical controls (no injury); surgical controls (injury without intervention); HA; AF; amniotic membrane suspended in phosphate buffered saline (AM); lyophilized hyaluronic acid reconstituted with amniotic fluid (HA + AF); and amniotic membrane suspended in hyaluronic acid (HA + AM). These groups were analyzed at three months and one year post-injury. Tissue was analyzed for these groups similar to Phase 1.

Results: For Phase 1 rheology, no significant differences were detected between control VFs at week four or week 10. Mean moduli values for control VFs were higher than treated VFs at week four, but not week 10. Treated VFs were comparable across groups at week four, but there was an intervention effect at week 10, with the AF group exhibiting higher elastic modulus values (6150 Pa) than saline (3635 Pa) and HA (4277 Pa) groups. For PCR, the significant comparisons were all for IL-1 β as follows: AF control values were lower than AF treated values at week 10; saline control values at week four were higher than at week 10; AF control values at week four were higher than at week 10; and saline treated values at week four were lower than AF treated values at week four.

Studies for Phase 2 are ongoing, but preliminary results from rheological analysis of three months animals will be presented at the meeting.

Conclusions: Based on rheology results of Phase 1, AF treated VFs were different/stiffer than uninjured controls, saline, and HA groups at 10 weeks. HA was consistent across acute and chronic time points, while saline was the most variable. Where significance was detected, inflammatory marker levels trended as expected. These results suggest a potential for the use of AF as a treatment for VF scar, but results from Phase 2 will further inform this topic.

Histological Evaluation of Human Laryngeal Embryogenesis

Vlasta Lungova¹, Jessica M. Fernandez¹, Susan L. Thibeault¹

¹ Department of Surgery, University of Wisconsin, Madison, WI, USA

Keywords: laryngeal cartilage development, epithelial lamina recanalization, epithelial invagination and cavitation

Abstract

Introduction & rationale:

Understanding the morphogenesis of the human larynx holds significant clinical importance, especially for pediatricians. Symptoms of congenital laryngeal anomalies often present at birth but may be delayed. Clinical findings range from subtle changes in feeding and voice to life-threatening respiratory obstructions that require urgent airway interventions to restore respiratory function and voicing. There are often chronic problems with voicing that persist for years (Hartnick & Robin, 2000).

The human larynx begins its development around gestational week 4 (GW 4), corresponding to days 26–28 post-conception (PC), in the region between the fourth pharyngeal pouch and the cranial portion of the respiratory diverticulum, which emerges as an outgrowth of the anterior foregut (Zaw-Tun & Burdi, 1985). This area is referred to as the primitive laryngopharynx (PLPh) (Zaw-Tun & Burdi, 1985). By GW 5 (days 30–35 PC), the PLPh elongates and becomes bilaterally compressed, leading to the medial approximation and fusion of lateral walls, forming the epithelial lamina (EL) (Rucci, Romagnoli, Casucci, & Ferlito, 2004). The EL temporarily occludes the laryngeal lumen and gradually disintegrates between GW 6 and 12 (days 42–84 PC). As a result, the lateral walls with the prospective vocal folds (VF) move apart, opening the laryngotracheal tube (Zaw-Tun & Burdi, 1985). Laryngeal and vocal fold (VF) structures continue to develop and differentiate. By GW 23, VF are fully functional, capable of opening and closing the glottis (Liberty et al., 2013). Common congenital laryngeal defects stem from incomplete EL recanalization, abnormal cartilage development, and neuromuscular immaturity (Hartnick & Robin, 2000). Most current knowledge about human laryngeal embryogenesis comes from mouse models, (Lungova, Verheyden, Herriges, Sun, & Thibeault, 2015), as detailed understanding of human development remains very limited.

Objectives:

This study provides a histological evaluation of human laryngeal morphogenesis during the early and mid-developmental stages, with a focus on cartilage development, epithelial differentiation, and recanalization processes.

Methods: Human embryonic/fetal larynges at days 52, 67, 76, and 96 post-conceptions (PC) were obtained from the Birth Defects Research Laboratory, University of Washington, with approval from the Stem Cell Research Oversight (SCRO) Committee, UW Madison. The larynges were fixed in 4% paraformaldehyde overnight, dehydrated in ethanol, treated with xylene, embedded in paraffin, and sectioned at 5 µm. For histology, sections were deparaffinized, rehydrated, and stained with hematoxylin and eosin (HE) to assess morphology and safranin-o-stain to visualize developing cartilages. To label epithelial layers, we performed double immunofluorescent staining for p63 and cytokeratin (K) 14, markers of the early stratified epithelium (Lungova, Verheyden, Herriges, Sun, & Thibeault, 2015).

Results:

At day 52 PC, the EL is in the early stages of recanalization (Figure 1A). Thyroid (TC) and cricoid cartilages (CC) are clearly defined, while the arytenoid cartilages (AC) and epiglottis (Epi) begin to condense at their prospective sites (Figure 1B). K14 and p63-positive epithelial cells line the expanding laryngeal cavities and the EL (Figure 1C).

By days 67 and 76 PC, the larynx expands, and the laryngeal lumen continues recanalization (Figure 1D, E, G, and H), with two remaining constrictions: one at the level of the VF and another at the caudal portion of the CC and first tracheal ring (Figure 1D, F). Failure of recanalization at these sites can lead to laryngeal webs or subglottic stenosis respectively (Hartnick & Robin, 2000). During recanalization, we also observed some epithelial cells delaminate from the surface epithelium, migrate in clusters deeper into the tissue, and expand to form cavities that eventually fuse with the lumen (Figure 1G–I). As for developing cartilages, both AC and Epi remain in mesenchymal condensation at day 67 (Figure 1E), with the AC becoming clearly defined by day 76 PC (Figure 1G, and H).

At day 96 PC, the larynx further expands, epithelial invagination and cavitation continues, likely contributing to the separation of the false vocal folds (FVF) from the true vocal folds (TVF) (Figure 1J–L). Epithelial clusters that remain deeper in the tissue differentiate into future glandular structures (Figure 1L). At day 96, both AC and Epi appear to be clearly distinguishable (Figure 1K).

Conclusions:

We provided a histological evaluation of normal laryngeal morphogenesis during early and mid-developmental stages. Our findings demonstrate that laryngeal cartilages develop in a sequential manner: the TC and CC complex forms first to provide structural support, while the AC and Epi develop later, when the larynx expands to accommodate both structures. Additionally, we showed that the epithelium likely plays an active role in recanalization through invagination

and cavitation, which may accelerate this process. The epithelium also appears to contribute to the separation of the true vocal folds from the vestibular folds (false VF).

Acknowledgement:

This work was funded by NIDCD R01020734.

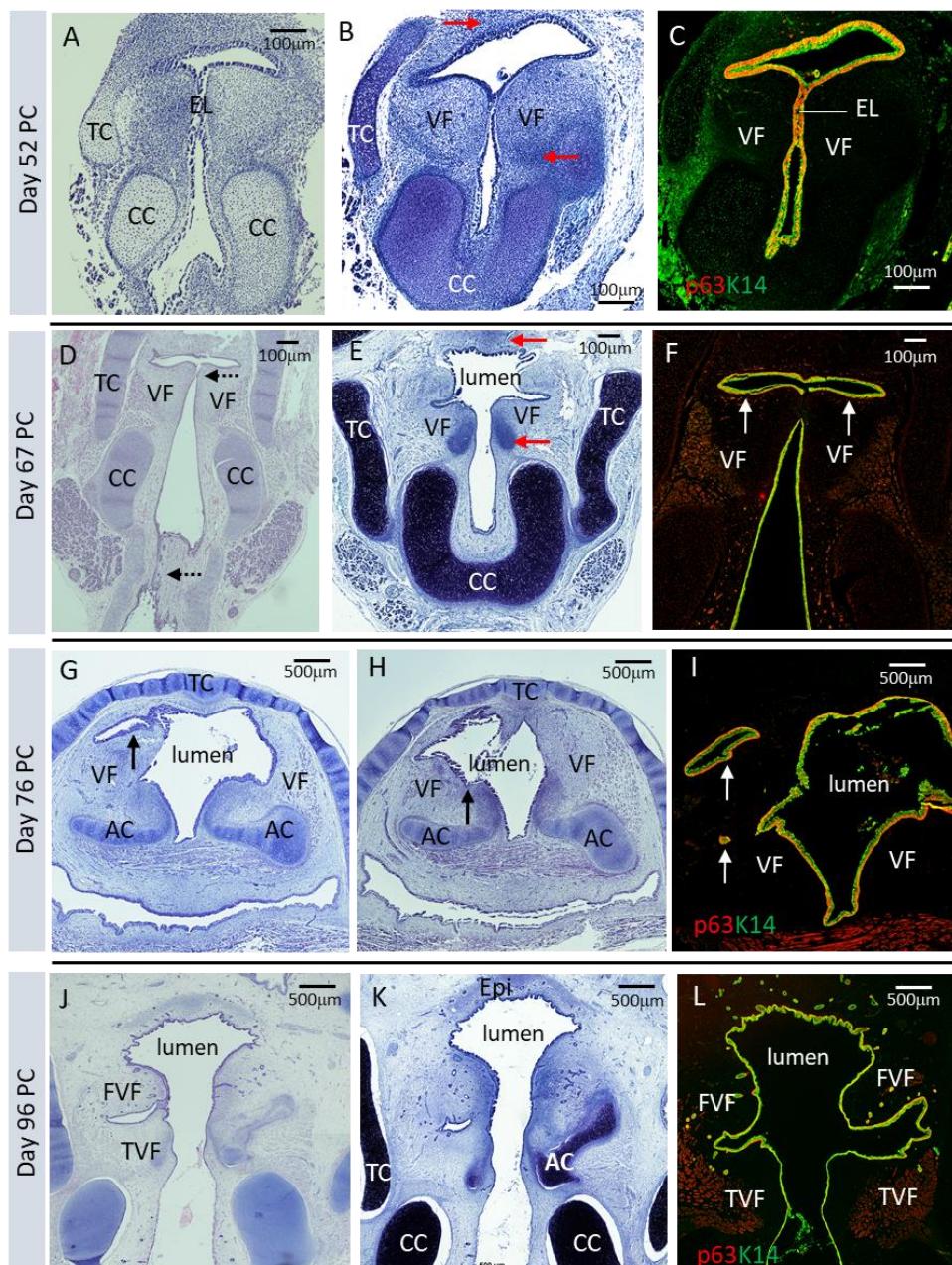


Figure 1: Morphological changes in developing human larynx. Hematoxylin-eosin staining of human larynx at day 52 post-conception, coronal sections (A). Safranin-O staining showing developing laryngeal cartilages (B). Anti-p63 (red) and anti-cytokeratin 14 (green) staining showing early stratified epithelium (C). Hematoxylin-eosin staining of human larynx at day 67 post-conception, coronal sections (D). Safranin-O staining showing developing laryngeal cartilages (E). Anti-p63 (red) and anti-cytokeratin 14 (green) staining showing early stratified epithelium lining the laryngeal cavities and vocal folds (F). Hematoxylin-eosin staining of human larynx at day 76 post-conception, transversal sections (G, H). Anti-p63 (red) and anti-cytokeratin 14 (green) staining showing early stratified epithelium lining the laryngeal lumen and the cavities (I). Hematoxylin-eosin staining of human larynx at day 96 post-conception, coronal sections (J). Safranin-O staining showing laryngeal cartilages (K). Anti-p63 (red) and anti-cytokeratin 14 (green) staining showing early stratified epithelium lining the lumen (L). Red solid arrows at the panels of B, E denote the mesenchymal condensation sites for the arytenoid cartilages and epiglottis. Black dashed arrows in the panel of D denote two constriction sites at the level of the vocal folds and the subglottis. Black solid arrows in the panels of G, H and white solid arrows in the panels of F, I denote epithelial clusters and expanding cavities. **Abbreviations:** AC, arytenoid cartilages; CC, cricoid cartilage; Epi, epiglottis; FVF, false vocal folds; TC, thyroid cartilage; TVF, true vocal folds; VF, vocal folds.

References:

- Hartnick, C. J., & Robin, C. (2000). Congenital Laryngeal Anomalies Laryngeal Atresia, Stenosis, Webs, and Clefts. *Otolaryngologic Clinics of North America*, 33(6), 1–10.
- Liberty, G., Boldes, R., Shen, O., Shaul, C., Cohen, S. M., & Yagel, S. (2013). The fetal larynx and pharynx: Structure and development on two- and three-dimensional ultrasound. *Ultrasound in Obstetrics and Gynecology*, 42(2), 140–148. <https://doi.org/10.1002/uog.12358>
- Lungova, V., Verheyden, J. M., Herriges, J., Sun, X., & Thibeault, S. L. (2015). Ontogeny of the mouse vocal fold epithelium. *Developmental Biology*, 399(2). <https://doi.org/10.1016/j.ydbio.2014.12.037>
- Rucci, L., Romagnoli, P., Casucci, A., & Ferlito, A. (2004). Embryological study of the glottic site and clinical implications. *Oral Oncology*, 40(10), 1017–1025. <https://doi.org/10.1016/j.oraloncology.2004.05.004>
- Zaw-Tun, H. A., & Burdi, A. R. (1985). Reexamination of the origin and early development of the human larynx. *Acta Anatomica*, 122(3), 163–184.

Laryngoresponders: are they more common in individuals diagnosed with vocal hyperfunction?

Daphne S. Toglia¹, Nicole Tomassi¹, Allison S. Aaron², Turley Duque², Kaitlyn A. Siedman², Jose M. Rojas², Lauren F. Tracy², J. Pieter Noordzij⁴, Cara E. Stepp^{1,2,3,4}

¹ Graduate Program for Neuroscience, Boston University, Boston, MA, 02215, USA

² Department of Speech, Language, and Hearing Sciences, Boston University, Boston, MA, 02215, USA

³ Department of Biomedical Engineering, Boston University, Boston, MA, 02215, USA

⁴ Department of Otolaryngology, Boston University School of Medicine, Boston, MA, 02215, USA

Keywords: Physical Report Form; Cepstral Peak Prominence; Low-to-High Spectral Energy Ratio

Abstract

Introduction & rationale: Vocal hyperfunction (VH), characterized by excessive perilyngeal muscle activity, is a key contributor to many common voice disorders (Roy et al., 2005). An explanation for why individuals develop VH has led to the concept of laryngoresponders (LRs) (Aronson, 1990). LRs were proposed as individuals with a psychological predisposition to hyperfunctional vocal behaviors, characterized by strained phonation or uncoordinated respiratory-phonatory behaviors (Becker et al., 2022). Recent studies piloted the Physical Report Form as a self-report tool adapted from chronic pain research to classify LRs based on body-mapping and laryngeal symptom descriptions (Becker, 2019; Becker et al., 2022). However, these studies assessed only individuals without VH, leaving the relationship between LRs and VH unexplored, as well as the potential influence stress has on this relationship. Stress has been shown to affect voice production by altering acoustic measures tied to autonomic arousal (Boyer et al., 2018; Kleinow & Smith, 2006; MacPherson, 2019; MacPherson et al., 2017), with one study reporting decreased low-to-high spectral energy ratio (L/H ratio) and increased cepstral peak prominence (CPP) (MacPherson et al., 2017). However, the effect of stress on acoustic correlates of voice quality in individuals with heightened laryngeal sensitivity, particularly in those who have vocal hyperfunction, remains unknown.

Objectives: The purpose of this study was two-fold. The first examined whether the incidence of LRs—identified using the Physical Report Form, a tool that assesses susceptibility to laryngeal sensitivity during high stress—differed between individuals with and without VH. We hypothesized that there would be a greater proportion of LRs in the VH group than in the control group. For the second, we investigated how LRs and non-laryngoresponders (NLRs) responded to stress during voice production induced by the Stroop task, a cognitive task that involves naming the color of a word when the word itself may denote a different color (e.g., the word “red” written in blue ink). The incongruent condition (when the word and color mismatch) has been shown to induce cognitive stress, whereas the congruent condition (when they match) does not (Abur et al., 2023; MacPherson, 2019; MacPherson et al., 2017). We hypothesized that all speakers would exhibit decreased L/H ratios and increased CPP during the incongruent condition relative to the congruent condition and that these effects would be more pronounced in LRs compared to NLRs.

Methods:

Participants. A total of 27 individuals with VH (21 females, 6 males; $M = 30.4$, $SD = 11.7$ years) and 27 age- and sex-matched controls without voice disorders ($M = 29.0$, $SD = 12.6$ years) participated with informed consent under Boston University IRB approval. All participants were fluent speakers of English with no recent voice/speech therapy (past 6 months) or other relevant comorbidities. Voice disorder diagnoses (with and without phonotrauma) were made by board-certified, fellowship-trained laryngologists at either Boston Medical Center or the Massachusetts General Hospital Voice Center.

Identifying Laryngoresponders and Non-Laryngoresponders. After acoustic data collection, participants completed the Physical Report Form. Experimenters instructed participants to mark areas on a body diagram that they perceived as chronically sensitive or reactive to stress. After data collection, a trained technician assessed the Physical Report Form and classified participants as LRs if they (a) identified the front-of-neck region as vulnerable to stress and (b) described symptoms suggestive of laryngeal involvement (e.g., globus sensation, throat tightness) (Becker, 2019; Becker et al., 2022). Others were categorized as NLRs. To ensure reliability in classification, an independent rater assessed a randomly selected subset of participants ($N = 14$), and inter-rater agreement was calculated using Cohen's Kappa. The analysis yielded a κ value of 1.00, indicating complete agreement between raters.

Procedure. Acoustic recordings of each participant were acquired using a head-mounted microphone (Shure SM35XLR) placed 45° from the midline and 7 cm from the lips. These were recorded with 16-bit resolution and at a sampling rate of 44.1 kHz in a sound-treated room at Boston University. Participants read a total of 12 sentences containing an embedded Stroop task. Six sentences formed the congruent condition, while the other six formed the incongruent condition. The congruent condition always preceded incongruent due to the duration of laryngeal responses to cognitive stress. This response may last for several minutes (Helou et al., 2013), and thus, a fixed order ensured that

acoustic measures from the congruent condition were not contaminated by stress induced during the incongruent condition.

Voice Analysis. Recordings were preprocessed using a customized MATLAB script to remove pauses (v6.0.50). Cepstral peak prominence (CPP) and low-to-high spectral energy ratio (L/H ratio) were extracted using a customized Praat script. Briefly, CPP was computed from a PowerCepstrogram with a 60 ms analysis window, 0.002 s time step, and a frequency range up to 5000 Hz, with parameters set for robust estimation. The L/H ratio was obtained from the spectral energy difference between 0–4000 Hz (low-frequency band) and 4000–20,000 Hz (high-frequency band).

Statistical Analysis. A Fisher's exact test was used to examine the association between LR status and the presence of VH. Two mixed-effects models were conducted to assess the effects of Stroop condition (incongruent vs. congruent), group (VH vs. control) and LR status (LR vs. NLR) on CPP and L/H Ratio, including interaction effects.

Results: A Fisher's exact test revealed a positive association between LR status ($p = 0.027$) and the presence of VH, such that people who were identified as LR were more likely to be within the VH group than the control group. A mixed-effects model revealed significant main effects of Stroop condition ($p = 0.019$) and LR status ($p = 0.040$) on CPP. However, no interaction effects were statistically significant ($p > 0.05$). A second mixed-effects model for the L/H Ratio revealed no statistically significant main or interaction effects ($p > 0.05$).

Conclusions: Our results support the hypothesis that LR status is related to the presence of VH. These findings suggest that LR status could be a predictive marker for individuals at risk of developing VH or that individuals with VH are more aware of their laryngeal sensations than those who do not have VH. Incorporating assessments for LR traits into clinical evaluations could enhance early detection and targeted intervention. The mixed-effects models revealed that CPP values were significantly influenced by LR status and Stroop task condition, but there was no significant interaction between these variables. LRs exhibited lower CPP values, suggesting that they spoke with a more dysphonic voice quality overall. CPP values increased under the incongruent condition for all speakers, indicating use of a more pressed voice during stress. However, the degree of change in CPP between the congruent and incongruent conditions was not modulated by LR status: both LRs and non-LRs exhibited similar responses to the Stroop task. These results suggest that although LRs may have distinct voice characteristics, their vocal adjustments to acute stressors are comparable to those of NLRs.

References

- Abur, D., MacPherson, M. K., Shembel, A. C., & Stepp, C. E. (2023). Acoustic Measures of Voice and Physiologic Measures of Autonomic Arousal During Speech as a Function of Cognitive Load in Older Adults. *Journal of Voice*, 37(2), 194–202.
- Aronson, A. E. (1990). Clinical voice disorders. *An Interdisciplinary Approach*, 157–197.
- Becker, D. R. (2019, June 4). *Vocal Manifestations of Reported Past Trauma* [University of Pittsburgh ETD]. University of Pittsburgh.
- Becker, D. R., Welch, B., Monti, E., Sullivan, H., & Helou, L. B. (2022). Investigating Past Trauma in Laryngoresponders Versus Non-Laryngoresponders: Piloting New Methods in an Exploratory Study. *Journal of Voice*.
- Boyer, S., Paubel, P.-V., Ruiz, R., El, Y. R., & Daurat, A. (2018). Human Voice as a Measure of Mental Load Level. *Journal of Speech, Language, and Hearing Research*, 61(11), 2722–2734.
- Kleinow, J., & Smith, A. (2006). Potential interactions among linguistic, autonomic, and motor factors in speech. *Developmental Psychobiology*, 48(4), 275–287.
- Helou, L. B., Wang, W., Ashmore, R. C., Rosen, C. A., & Abbott, K. V. (2013). Intrinsic laryngeal muscle activity in response to autonomic nervous system activation. *The Laryngoscope*, 123(11), 2756–2765.
- MacPherson, M. K. (2019). Cognitive Load Affects Speech Motor Performance Differently in Older and Younger Adults. *Journal of Speech, Language, and Hearing Research*, 62(5), 1258–1277.
- MacPherson, M. K., Abur, D., & Stepp, C. E. (2017). Acoustic Measures of Voice and Physiologic Measures of Autonomic Arousal during Speech as a Function of Cognitive Load. *Journal of Voice: Official Journal of the Voice Foundation*, 31(4), 504.e1-504.e9.
- Roy, N., Merrill, R. M., Gray, S. D., & Smith, E. M. (2005). Voice Disorders in the General Population: Prevalence, Risk Factors, and Occupational Impact. *The Laryngoscope*, 115(11), 1988–1995.

Post-Influenza Airway Injury: Immune Cell Recruitment and Oncogene Myc-driven Basal Cell Regeneration

Alexander G. Foote, PhD¹, Xin Sun, PhD^{1,2}

¹ Department of Pediatrics, University of California – San Diego, San Diego, CA, USA

² Division of Biological Sciences, University of California – San Diego, San Diego, CA, USA

Keywords: upper respiratory epithelium, influenza-viral injury, immunity

Abstract

Introduction & rationale: Influenza A virus (IAV) is rapidly detected in the upper respiratory tract by epithelial and immune cells, triggering antiviral inflammatory responses.¹ Inflammation of the laryngeal and proximal airway mucosa, commonly associated with acute and chronic laryngitis, plays a critical role in regulating local immune responses.² Viral laryngitis can manifest in a range of clinical conditions, including chronic cough, idiopathic ulcerative laryngitis, and post-viral vagal neuropathy.^{3,4} While IAV is a well-established respiratory pathogen responsible for significant global morbidity, its effects on the larynx and surrounding tissues remain poorly understood.

Objectives: This study aims to elucidate epithelial repair mechanisms and the complex interplay between innate and adaptive immune responses in the larynx and upper respiratory tract following influenza infection.

Methods: We employed a combination of mouse genetics, cell-based assays, and single-cell transcriptomics to investigate epithelial remodeling and immune cell dynamics at acute (7 days post-infection, 7DP-IAV) and chronic (21DP-IAV) phases in wild-type mice. To assess the functional role of *Myc*, a proto-oncogene transcription factor, in airway basal cell-mediated repair, we utilized *Krt5creER;Myc* mutant mice to ablate *Myc* expression in basal cells following IAV-induced injury.

Results: IAV infection induced an acute influx of Ly6G+ neutrophils and a sustained presence of CD8+ NKT-like cells within the epithelial compartment of the subglottis and proximal trachea. Notably, we identified a cycling-repair epithelial cell population expressing ectopic *Myc*, which emerged in response to epithelial injury. Functional assays demonstrated that *Myc* plays a critical role in promoting KRT5+ basal cell-driven proliferative repair following IAV infection.

Conclusions: Our findings characterize distinct epithelial and immune cell populations that contribute to host immunity and tissue repair following viral infection. We establish a key role for *Myc* in regulating basal cell-mediated epithelial regeneration. Moreover, in the chronic disease setting, we identify a cytotoxic lymphocyte population with a gene signature indicative of CD8+ NKT-like cells. Future studies will determine whether this unique immune population contributes to immunologic memory against secondary insults or exerts deleterious effects on surrounding tissues.

References:

1. Chen, X., Liu, S., Goraya, M. U., Maarouf, M., Huang, S., & Chen, J. L. (2018). Host immune response to influenza A virus infection. *Frontiers in immunology*, 9, 320.
2. Jetté, M. (2016). Toward an Understanding of the Pathophysiology of Chronic Laryngitis. *Perspectives of the ASHA special interest groups*, 1(3), 14-25.
3. Dominguez, L. M., & Simpson, C. B. (2015). Viral laryngitis: a mimic and a monster—range, presentation, management. *Current Opinion in Otolaryngology & Head and Neck Surgery*, 23(6), 454-458.
4. Jaworek, A. J., Earasi, K., Lyons, K. M., Daggumati, S., Hu, A., & Sataloff, R. T. (2018). Acute infectious laryngitis: A case series. *Ear, Nose & Throat Journal*, 97(9), 306-313.

Comparability and quantification of vocal tract-related efficiency strategies of different voice qualities in four professional singers

Fiona Stritt^{1,2}, Stefanie Rummel³, Johannes Fischer^{2,4}, Michael Bock^{2,4}, Bernhard Richter^{1,2}, Matthias Echternach⁵, Louisa Traser^{1,2*}

¹ Freiburg Institute for Musicians' Medicine, Medical Center – University of Freiburg, Faculty of Medicine, Germany

² Faculty of Medicine, University of Freiburg, Germany

³ Institut Rummel, Frankfurt, Germany

⁴ Department of Radiology, Medical Physics, Medical Center – University of Freiburg, Germany

⁵ Department of Otorhinolaryngology, Ludwig-Maximilians-Universität München, Division of Phoniatrics and Pediatric Audiology, LMU Klinikum, Munich, Germany

Keywords: 3D vocal tract MRI, finite-element-modeling, vocal tract efficiency

Introduction & rationale: Singing can exhibit a wide dynamic range, with particularly high volume levels in Western classical singing and belting. Under these demanding conditions efficient voice production is required to generate high sound pressure level with minimal energy and biomechanical load to maintain vocal health. Inefficient strategies may increase biomechanical stress, particularly on the vocal folds leading to secondary organic changes (Hirosaki et al., 2021; Jiang & Titze, 1994). Vocal tract (VT) efficiency can be quantified using the transfer function or the inertance; however, a single variable would be advantageous to directly compare the efficiency of different VT configurations. In a pilot study, we introduced a metric to quantify acoustic sound intensity at the glottis which determines the energy required to excite various VT configurations (Fleischer et al., 2022).

Objectives: To assess whether the differences in VT efficiency measures for various voice qualities are due to individual vocal approaches or whether they are linked to the VT configurations trained within a vocal school.

Methods: Four professional singers (2♀: S1,S3; 2♂: S2,S4) with advanced Estill Voice Training® (EVT®) qualifications phonated [a:] in six voice qualities (Speech, Falsetto, Sob, Opera, Oral Twang, Belting) at ♀415Hz/♂207Hz during 3D magnetic resonance imaging (MRI) of the VT. From the 3D MRI data, VT configurations were analyzed using image segmentation(Traser et al., 2017). Using finite element modeling the volume velocity transfer function (VVTF) was calculated as outlined in (Birkholz et al., 2020). As an efficiency measure, sound intensity at the glottis was calculated from lip sound recordings and numerically derived VT transfer functions (Fleischer et al., 2022). Represented as infinite-response-filters, glottal particle velocity was obtained by deconvolving lip sound pressure with the VT impulse response, while glottal acoustic pressure was derived by filtering the particle velocity.

Results: VT configurations showed strong similarities across all singers and voice qualities. Belting and Twang featured hypopharyngeal narrowing, a trumpet-shaped oropharynx, and a reduced VT length ("megaphone" shape). Opera and Sob had an hourglass-shaped VT with hypopharyngeal widening, mid-pharyngeal narrowing, moderate oral cavity widening, and a low larynx—most pronounced in Sob. Falsetto and Speech exhibited intermediate VT lengths and the greatest individual variation.

The VVTF for Opera showed a formant cluster peaking between 3–4 kHz for all singers (Fig.1), while Sob exhibited anti-resonances in this range. Overall, the megaphone VT group had minimal energy loss, though individual anti-resonances were noted. Falsetto and Speech displayed neutral transmission in S1, resembled Opera in S2, and aligned more with Twang and Belting in S3 and S4. Efficiency calculations mostly indicated that Belting showed an overall higher efficiency than Oral Twang, whereas Sob required greater glottal energy to excite the VT than Opera. Individual differences were observed in the excitation of Falsetto, Speech.

Discussion and Conclusions:

VT configurations showed distinct patterns for each voice quality with high inter-individual consistency. A grouping considering the configurations into megaphone- (Belting/Twang), neutral- (Speech/ Falsetto) (Titze et al., 2017), or hourglass- (Opera/Sob), is also reflected in the acoustic properties.

A reinforcement of harmonic energy in the psychoacoustically relevant range of 2–4 kHz was pronounced in Opera, Belting and Twang. While Opera exhibits the singer's formant cluster, Belting and Oral Twang showed a general increase in overall energy levels. Anti-resonances in Sob may result from dependent side spaces in the hypopharynx (Feng & Howard, 2023; Vampola et al., 2015), with similar effects potentially explaining anti-resonances in Twang and Belting.

For Falsetto/Speech, inter-individual VVTF varied. EVT® distinguishes these qualities by vocal fold vibration patterns (Falsetto: breathy phonation; Speech: complete glottal closure with high vertical thickness), suggesting that VT shaping plays a secondary role, likely influenced by the singer's preferred and most frequently trained style.

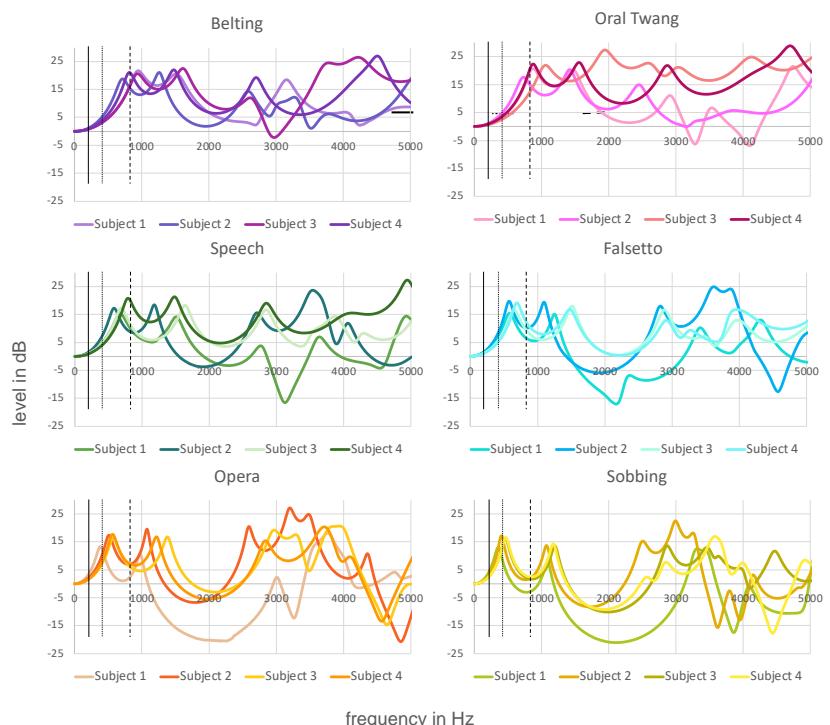


Fig. 1: Volume-velocity transfer functions for all subjects and voice qualities.
Vertical black lines representing the fundamental frequency and the first partial for both male and female singers. — 207 Hz 415 Hz – 830 Hz.

Preliminary analysis of the calculated glottal sound

intensity aligns with qualitative observations, showing that different VT configurations lead to varying energy levels. For example, Sob phonation demands significantly more glottal intensity than Opera to achieve the same output. Although the small sample size limits generalizability, the consistency in VT configurations and their acoustic properties among singers from the same vocal school suggests that it may be possible to apply defined resonance strategies with specific efficiency properties through training. Variations appear linked to more frequently practiced styles. Further research should explore how anatomical features affect ease of producing certain qualities. A single metric to quantify VT efficiency could enhance understanding and encourage conscious use of less efficient voice qualities, potentially preventing voice disorders when combined with acoustic amplification.

References:

- Birkholz, P., Kürbis, S., Stone, S., Häsner, P., Blandin, R., & Fleischer, M. (2020). Printable 3D vocal tract shapes from MRI data and their acoustic and aerodynamic properties. *Scientific Data*, 7(1).
- Feng, M., & Howard, D. M. (2023). The Dynamic Effect of the Valleculeae on Singing Voice – An Exploratory Study Using 3D Printed Vocal Tracts. *Journal of Voice*, 37(2), 178–186.
- Fleischer, M., Rummel, S., Stritt, F., Fischer, J., Bock, M., Echternach, M., Richter, B., & Traser, L. (2022). Voice efficiency for different voice qualities combining experimentally derived sound signals and numerical modeling of the vocal tract. *Frontiers in Physiology*, 13.
- Hirosaki, M., Kanazawa, T., Komazawa, D., Konomi, U., Sakaguchi, Y., Katori, Y., & Watanabe, Y. (2021). Predominant Vertical Location of Benign Vocal Fold Lesions by Sex and Music Genre: Implication for Pathogenesis. *The Laryngoscope*, 131.
- Jiang, J. J., & Titze, I. R. (1994). Measurement of vocal fold intraglottal pressure and impact stress. *Journal of Voice*, 8(2), 132–144.
- Steinhauer, K., McDonald Klimek, M., & Estill, J. (2017). *The Estill Voice Model: Theory & Translation*. Estill Voice International.
- Titze, I. R., Maxfield, L. M., & Walker, M. C. (2017). A Formant Range Profile for Singers. *Journal of Voice*, 31(3), 382.e9–382.e13.
- Vampola, T., Horáček, J., & Svec, J. (2015). Modeling the Influence of Piriform Sinuses and Valleculeae on the Vocal Tract Resonances and Antiresonances. *Acta Acustica United with Acustica*, 101.
- Traser, L., Birkholz, P., Viktoria Flügge, T., Kammerer, R., Burdumy, M., Richter, B., Gerrit Korvink, J., & Echternach, M. (2017). Relevance of the Implementation of Teeth in Three-Dimensional Vocal Tract Models. *Journal of Speech, Language, and Hearing Research*, 1–15.

Vowel onset VRT: a comparison of signal envelope algorithms and their key parameters

John Holik¹, Duy Duong Nguyen¹, Tünde Szalay¹, Tomás Arias-Vergara², Michael Döllinger³, Catherine Madill¹

¹ Sydney Voice Lab (Dr Liang Voice Program), Discipline of Speech Pathology, University of Sydney, Australia.

² Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Erlangen, Germany

³ Division of Phoniatrics and Paediatric Audiology at the Department of Otorhinolaryngology Head & Neck Surgery, University Hospital Erlangen, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Erlangen, Germany

Keywords: *Voice Onset, Vocal Rise Time, Signal Processing*

Introduction & rationale:

Vocal Rise Time (VRT) is the measurement of the time taken for the voice to reach a steady state when a vowel is produced from rest (Koike, 1967). VRT is a fundamental feature of the onset of the vowel driven voice and no agreed upon standards exist to measure this feature (Chacon et al., 2024). Measurement of this phenomena requires deciding when the sound begins and when the vowel reaches a steady state. This is most conveniently achieved by examining an envelope of the sound waveform and defining vowel onset and steady state onset threshold parameters. This has previously been done manually (Maryn & Poncelet, 2021), which is time and labour intensive, thus not practical in clinical settings. To reduce the time-demand of VRT analysis, semi-automated methods of estimating VRT were examined using multiple algorithms and parameters.

The Voice Onset Analysis Tool (VOAT) (Arias-Vergara et al., 2024) is a Python program built to measure VRT manually and in batches. Three algorithms: Hilbert, Peak Amplitudes and Root Mean Square (RMS) are used to operate on the raw signal and generate an envelope. Voice Analysis Studio (VAS, Sydney Voice Lab) is an in-house MATLAB program with a manual VRT option and uses the RMS algorithm. The Hilbert algorithm finds the instantaneous amplitude of the signal and then smooths this by convolving it with a Gaussian Window of varying width (see Figure 1.). The Peak Amplitudes algorithm finds consecutive signal peaks of varying minimum distance from each other and generates an envelope by interpolating between these peaks. Lastly, the Root Mean Square of a varying window width is calculated and then shifted by a set time frame, generating an envelope from these values.

A set of varying parameters, unique to each algorithm were used to take multiple measures of the same audio, producing signal envelopes to measure VRT. The difference between the envelope and the signal trace, the Area Under the Envelope (AUE), was used as a proxy for a visual ‘closeness of fit’. A combination of the AUE and the standard deviation of the VRT values was used to find an optimal envelope varying parameter combination to be used by each algorithm for measurement.

Objectives:

We aimed to examine the effect of varying parameter combinations on VRT measurements to establish which parameters produce reliable replicable results.

Methods:

Two VRT programs with three different algorithmic approaches were used to measure a ‘Normative’ (Norm) data set and an instructed ‘Laryngeal Manipulation’ (LM) data set with 114 and 9 non-disordered participants respectively producing the sound /a/.

Each algorithm was executed with six different parameter values:

- Hilbert used Gaussian Convolution windows of width 0, 40, 80, 120, 160, 200 ms.
- Peak Amplitudes defined minimum distances between peaks of 0, 1, 2, 3, 4, 5 smoothing factor (SF).
- RMS defined window calculation width of 0, 20, 40, 60, 80, 100 ms.

The thresholds for calculating the VRT onset are 10% and VRT offset at 90% of maximum signal.

Table 1 - VRT averages over all varying parameters for initial /a/ [ms]

Source	Algorithm	Mean	SD
Norm	Hilbert	57	38
LM	Hilbert	41	30
Norm	Peak Amp	57	36
LM	Peak Amp	70	32
Norm	RMS VOAT	65	37
LM	RMS VOAT	59	28
Norm	RMS VAS	66	29
LM	RMS VAS	42	25

Figure 1 - Hilbert transform with Gaussian Convolution window

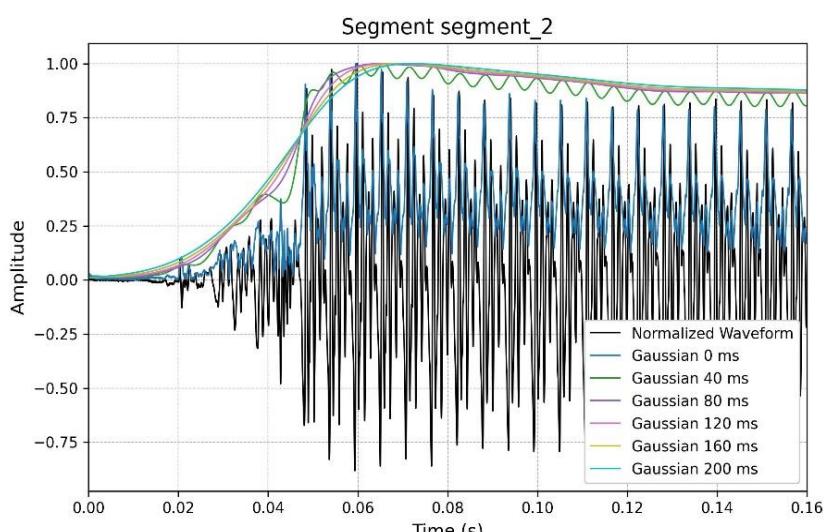


Figure 2 - Normative VRT values for all algorithms and Varying Parameters [ms]

Results:

Figure 1 shows an example of a signal and its Hilbert transform. The Gaussian Convolution window of 0 ms functions as an absolute value of the signal, this allows for its use as the signal trace when calculating the AUE. Shorter convolution window settings impart more details of the signal to the envelope. Excessive smoothing is apparent with windows over 40 ms, resulting in inflated VRT values. The 40 ms window envelope has the smallest AUE, as can be seen in Figure 1 – this is the closest envelope to the signal trace.

Table 1 shows the VRT values averaged across all varying parameters for the /a/ sound. The Normative RMS values show close agreement, but on inspection of the scatter plot in Figure 2, the spread of values converges as the window length increases. This is a false increase in accuracy as the envelopes are simply smoother, reducing spread in values. The Hilbert and Peak Amplitudes Normative values agree and are smaller than the RMS values, yet they are all within the same SD range.

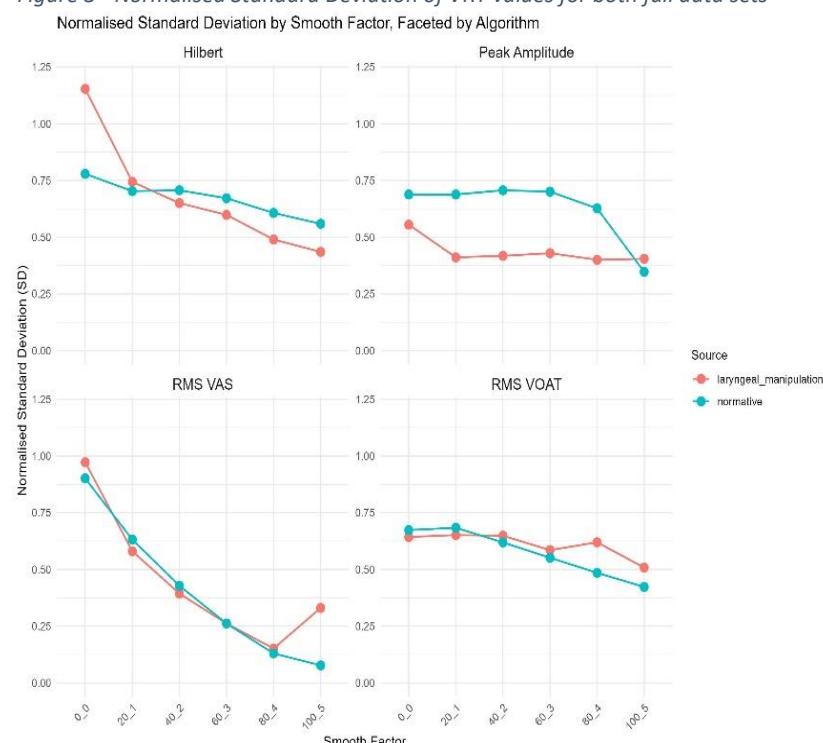
Figure 3 presents the normalized Standard Deviation of both full data sets over all varying parameters. For the Normative data set in both Hilbert and Peak Amplitude algorithms, the 20_1 parameter (Gaussian 40 ms, SF 1) is optimal, when the excessive smoothness of the largest 3 parameters are considered. The two RMS algorithms have different window shifting parameters, the VAS being much higher in resolution. This improving SD as the windows increase reflects the larger averaging effect over a longer window and is not an indication of increased precision.

Conclusions:

All envelope generating algorithms exhibit over-smoothing for the highest three varying parameters (windows or smoothing factor) and under smoothing for the lowest value. The optimal parameter is the second window/smoothing factor (Hilbert 40 ms, Peak Amplitude SF = 2 and RMS 20 ms). The Peak Amplitude algorithm produces results in agreement with the Hilbert algorithm, yet on individual inspection of envelopes, these are often not acceptable by the visual ‘closeness of fit’ criteria and require further investigation.

This work demonstrates that the VRT of vowel onsets can be measured using algorithm-generated envelopes, with the AUE ‘closeness of fit’ estimate serving as a proxy for a human eye assessment of the signal and its corresponding amplitude envelope. Our results will inform the use of VRT in clinical settings.

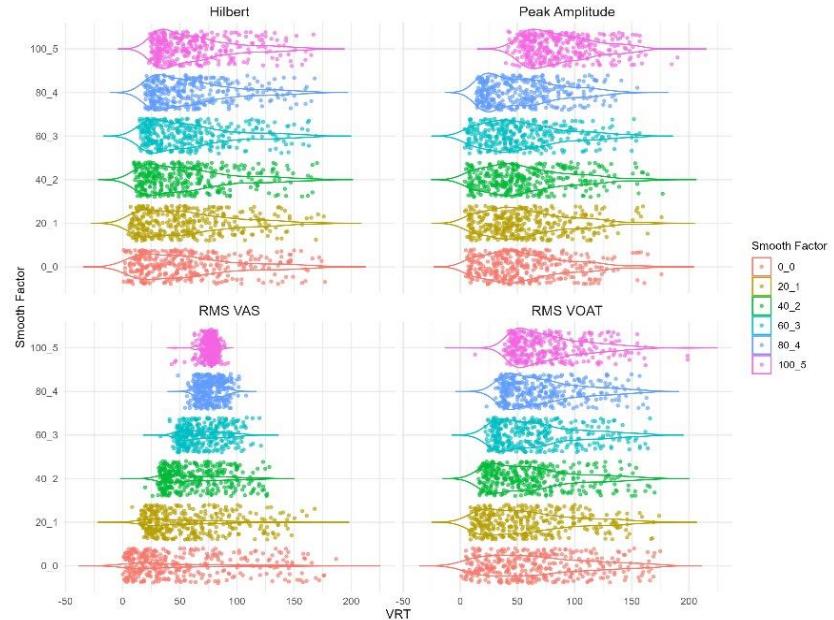
Figure 3 - Normalised Standard Deviation of VRT values for both full data sets



References:

- Arias-Vergara, T., Madill, C., Nguyen, D., Holik, J., and Döllinger, M. (2024). “VOAT: Voice Onset Analysis Tool,” *SoftwareX* 27, 101802. <https://doi.org/10.1016/j.softx.2024.101802>
- Chacon, A. M., Nguyen, D. D., Holik, J., Döllinger, M., Arias- Vergara, T., and Madill, C. J. (2024). “Vowel onset measures and their reliability, sensitivity and specificity: A systematic literature review,” *PLOS ONE* 19(5), e0301786. <https://doi.org/10.1371/journal.pone.0301786>
- Koike, Y. (1967). “Experimental Studies on Vocal Attack,” *Practica Oto-Rhino-Laryngologica* 60(8), 663–688. <https://doi.org/10.5631/jibirin.60.663>
- Maryn, Y, Poncelet, (2021). “How Reliable Is the Auditory-Perceptual Evaluation of Phonation Onset Hardness?” *Journal of Voice*, Vol.35 (6), p.869-875. <https://doi.org/10.1016/j.jvoice.2020.04.006>

Violin Plot of Normative data - Faceted by Algorithm



Pitch Reflex Responses in Individuals with Hearing Impairment via Acoustics and Electroencephalography

Tony Schelhorn¹, Benjamin Peschel², Michael Döllinger², Defne Abur^{3,4}, Ulrich Hoppe¹

¹ Division of Audiology, Department of Otorhinolaryngology, Head and Neck Surgery, University Hospital Erlangen, FAU Erlangen-Nürnberg, Germany

² Division of Phoniatrics and Pediatric Audiology, Department of Otorhinolaryngology, Head and Neck Surgery, University Hospital Erlangen, FAU Erlangen-Nürnberg, Germany

³ Department of Computational Linguistics, Centre for Language and Cognition Groningen, University of Groningen, The Netherlands

⁴ Research School of Behavioral and Cognitive Neurosciences, University of Groningen, Groningen, the Netherlands

Keywords: Auditory Perturbation, Hearing Impairment, Acoustic Change Complex

Abstract

Introduction & rationale:

In pitch shift experiments, participants receive auditory feedback of their own voice via headphones, which is then unexpectedly shifted in pitch. This triggers the pitch shift reflex (PSR) in which participants with typical hearing (TH) respond to the auditory perturbation by changing their own vocal pitch, which can be measured acoustically. Observing this response can give insights on the importance and function of auditory feedback for speech motor control. In the case of participants with hearing impairments (HI), their diminished auditory feedback increases both the potential information gain about the speech motor control complex and the difficulty of ensuring perceivable and comparable stimuli. By recording EEG signals during such experiments, the presence of the acoustic change complex (ACC) confirms activity in the auditory cortex in response to the given stimulus.

Objectives:

The current questions to be answered are (1) “Can hearing impaired participants perceive pitch shift stimuli given in the current paradigm and setup?” and (2) “How do HI and NH participants differ in their EEG response to the pitch shift?”.

Methods:

The current and preliminary analysis includes 7 participants with HI (5 female; 2 male; median age: 81) and 7 with TH (4 female; 3 male; median age: 67) without voice pathologies. The hearing impairments were confirmed to be sensorineural in nature and within a level of 40-80 dB_{HL} average hearing loss (better ear) during extensive pre-assessments. For the TH group, only participants with hearing loss below 20 dB_{HL} were included. All participants were screened for voice pathologies as part of the pre-assessments.

Participants were seated in a sound-attenuated room and given the task to imitate and hold a model /a/ phonation. The model phonation and all following feedback was provided through ER-2 earphones (Etymotic Research, Elk Grove Village, IL, USA). The model phonation ended after 1 s, after which the participants voice is recorded by an earset microphone, amplified and played back to the participants with the addition of pink noise. The delay between recording and auditory feedback was kept between 24-32 ms. Participants were instructed to aim for a volume of ca. 70 dB_{SPL} as indicated by a scale on a monitor 2 m away. This also provided a steady point of focus which reduces the influence of eye movement artifacts and reduces variations in overall vigilance.

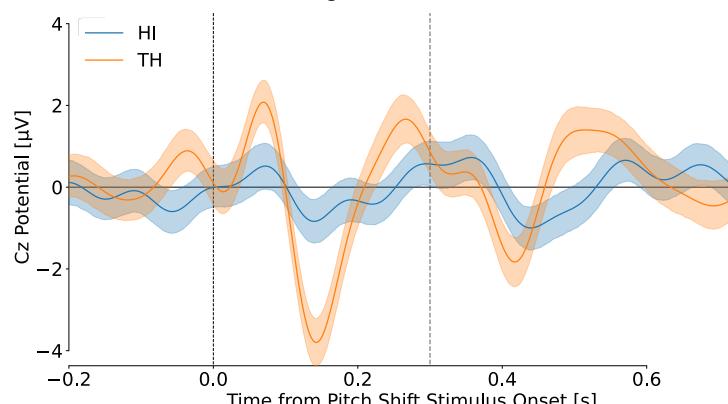


Figure 1: Average Cz potential over all trials from all participants of each group (Blue = hearing impaired; Orange = typically hearing). A second ACC response is visible after 300 ms caused by the return to un-shifted feedback.

At some point during each trial, randomly jittered between 2.0 s and 2.3 s the participant's feedback is pitch shifted by +200 Cent for 300 ms as a method to elicit ACC and PSR. Every trial had a length of 3.5 s. A total of 70 trials were completed per participant.

Brain activity was recorded at a sampling rate of 4800 Hz with a 32-channel EEG system (g.GAMMAsys, g.tec, Schiedlberg, Austria) arranged according to the international 10-20 system. EEG data was processed and analysed using MNE-Python [Larson2024]. Pre-processing was done by applying a 2-10 Hz bandpass filter, rejecting channels with excessive impedance, removing eye movement artifacts by use of independent component analysis, followed by auto-rejection and interpolation of unusable trials as provided by MNE-Python. P1, N1 and P2 of the ACC were selected as the peaks in potential at electrode Cz between 50-100 ms, 100-200 ms and 200-300 ms after stimulus onset respectively for the average response of each participant.

Statistical analysis on ACC parameters was performed in IBM SPSS. Mann-Whitney-U-tests were performed to compare peak amplitudes and latencies of the ACC between HI and TH groups.

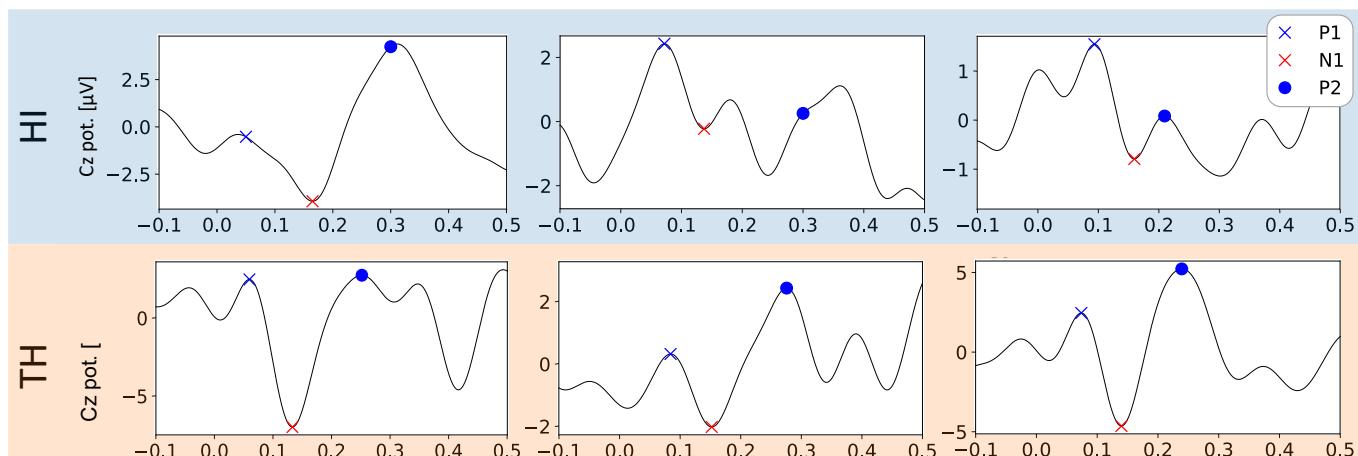


Figure 2: Example mean responses to pitch shifts measured at Cz for 3 participants with hearing impairment (HI; top-blue) and 3 participants with typical hearing (TH; bottom-orange). P1-N1-P2 detection is visualised as blue "x", red "x", and blue "o" respectively.

Results:

Only 3 participants with HI exhibited easily identifiable ACCs, while 4 participants did not show characteristic ACCs with corresponding components (see Figure 2). All ACCs from the TH group were detectable via the employed methods. Participants with TH showed larger P1 ($p=0.053$) and N1 ($p=0.007$) amplitudes than those with HI. The average P1 and N1 amplitudes at Cz measured $+2.4 (\pm 1.56)$ µV and $-4.2 (\pm 2.0)$ µV in the case of NH, and $+0.9 (\pm 1.0)$ µV and $-1.2 (\pm 1.5)$ µV for HI respectively. No significant difference in P1/N1/P2 latencies or P2 amplitude was found between the groups.

There was no correlation found between age and ACC parameters within the groups.

First analyses of acoustic data showed less pronounced PSR responses in participants with HI.

Conclusions:

HI participants show less pronounced ACCs when confronted with the current pitch shift stimulus. One possible cause is the diminished subjective volume of received auditory feedback, as it has been reported that decreased stimulus intensity generally results in lower N1 and P2 amplitudes as well as higher latencies [Adler1991]. More in-depth analyses on possible interactions between PSR and HI will follow as the study progresses.

References:

[Adler1991] - Adler, G., Adler, J. (1991). Auditory stimulus processing at different stimulus intensities as reflected by auditory evoked potentials. *Biological Psychiatry*. [https://doi.org/10.1016/0006-3223\(91\)90220-G](https://doi.org/10.1016/0006-3223(91)90220-G)

[Larson2024] - Larson, E., Gramfort, A., Engemann, D. A., Leppakangas, J., Brodbeck, C., Jas, M., Brooks, T. L., Sassenhagen, J., McCloy, D., Luessi, M., King, J.-R., Höchenberger, R., Goj, R., Brunner, C., Favelier, G., van Vliet, M., Wronkiewicz, M., Rockhill, A., Holdgraf, C., ... Iuzpaz. (2024). MNE-Python (v1.9.0). Zenodo. <https://doi.org/10.5281/zenodo.14519545>

Role of Piezo1 in Human Vocal Fold Fibroblasts

Se-in Kim^{1,2}, Vlasta Lungova², Susan Thibeault^{1,2}

¹ Department of Communication Sciences and Disorders, University of Wisconsin-Madison, Madison, Wisconsin, USA

² Department of Surgery, University of Wisconsin-Madison, Madison, Wisconsin, USA

Keywords: (3 maximum) human vocal fold fibroblast, pharmaceutical activation, Piezo1

Abstract

Introduction & rationale: Piezo is a lately discovered family of mechanosensitive cation channels consisting of Piezo1 and Piezo2 (Coste et al., 2010; Fang et al., 2021). Piezo channels are expressed in a wide range of mechanically sensitive tissues and cell types including but not limited to skin, lung, gastrointestinal tract, blood vessels, urinary tract, and neurons (Fang et al., 2021). In murine vocal folds, Piezo1 plays a role in murine vocal fold epithelial (re)modeling during development and after injury (Foote et al., 2022; Lungova et al., 2020).

Besides mechanical activation, Piezo channels can also be activated pharmaceutically by Yoda1. (Syeda et al., 2015). Yoda1 has been a convenient tool for understanding Piezo1's role because this synthetic molecule can activate Piezo1 without mechanical stimulation and lacks its effect on Piezo2 (Blythe et al., 2019; Lukacs et al., 2015; Syeda et al., 2015). Blythe et al., (2019) reported that Piezo1 activation of cardiac fibroblasts via Yoda1 led to pro-fibrotic IL-6 secretion via p38 MAPK pathway. Lukacs et al., (2015) used Yoda1 to discover a type of congenital lymphatic dysplasia associated with decreased Piezo1 function.

Human vocal folds (VF) are under significant phonatory stress. Human vocal fold fibroblasts (hVFF) are exposed to different magnitudes of physical forces based on the depth of the lamina propria, producing a gradient extracellular matrix structure (Gray, 2000). However, whether hVFF express mechanoproteins to sense these forces is not known. We hypothesized that hVFF are likely candidates for expressing Piezo1, given that this mechanosensitive channel is expressed in cardiac fibroblasts (Blythe et al., 2019) and that hVFF express multiple transcription factor genes critical to heart development (Foote et al., 2019). Next, we hypothesized that Yoda1 can activate Piezo1 channel in hVFF and prolonged (24-hour) pharmaceutical activation of the Piezo1 channel would alter the gene expression of hVFF.

Objectives: Investigate the expression of Piezo1 mechanosensitive channel and its biological role in in-vitro cultures of hVFF.

Methods: Immortalized hVFF were chosen for the present study for their relatively easier access and extended proliferation capacity compared to primary cell line (Chen & Thibeault, 2009). Four sex-balanced biological replicates of hVFF were cultured *in vitro*. Total protein was extracted to detect protein expression of Piezo1 using Western blot analysis. To demonstrate that the Piezo1 channel in hVFF was a functional channel, ratiometric fluorescence intracellular Ca²⁺ assay (Fura-2) was completed. Extracellular calcium influx through the Piezo1 channel was assessed when its agonist Yoda1 (10 μM) was added to the cell media. Extracellular calcium chelating conditions by pretreatment with EGTA were included to assess whether the response can be abolished.

Prior to initiating gene expression studies, a Lactate dehydrogenase (LDH)-Glo™ Cytotoxicity assay was completed to ensure that Yoda1 was not toxic to cells when exposed for 24 hours. This bioluminescence assay quantified LDH enzyme release to cell media, upon rupture of the cell membrane. Finally, to evaluate gene expression changes upon activation of Piezo1 channel, hVFF were incubated with Yoda1 (10 μM) or vehicle (DMSO1%) for 24 hours. After treatment, total RNA was extracted from the cells and underwent bulk RNA sequencing to discover Yoda1-mediated transcriptome changes in hVFF.

Results: Western blot for Piezo1 expressed protein bands immediately below 250kDa for all hVFF (Fig.1), corresponding with observed molecular weight range of Piezo1 protein (Ilkan et al., 2017). Results from intracellular Ca²⁺ assay (Fig. 2) demonstrated a significantly greater increase in intracellular Ca²⁺ concentration in response to Yoda1 than the vehicle injection, indicating that Piezo1 channels in hVFF can be opened by Yoda1 agonist with Ca²⁺ entry in the absence of mechanical stimulation (One-way ANOVA: $p=0.0019$, $F=13.648$; Tukey HSD test Yoda1 versus vehicle condition: $p=0.004124$). A 24-hour treatment of the hVFF with 10 μM Yoda1 was not significantly more cytotoxic than the vehicle as evidenced by comparable luminescence levels shown in LDH bioluminescent cytotoxicity studies (Fig.3) (One-way ANOVA; $p=0.90$, $F=0.19864$). Finally, transcriptomic data measuring 24-h Yoda1 treatment effect are presently being completed and will be done in time for presentation in June.

Conclusions: hVFF expressed the Piezo1 channel which can be opened with Yoda1 treatment. The effect of its pharmaceutical activation on hVFF gene expression is to be investigated.

Figure 1. Mechanosensitive channel expression in hVFF. Western blot detected Piezo1 protein (233kDa). Age and the sex of the hVFF cell donors listed at the top. Human Primary Small Airway Epithelial Cells (HSAEC) served as a positive control.

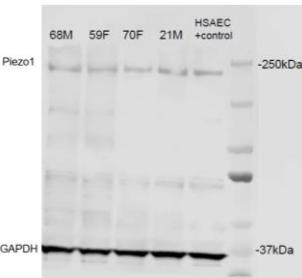


Figure 3. Percentage cytotoxicity of 24-hour Yoda1 treatment calculated based on luminescence of the released LDH relative to the max. LDH release control ($n=4$). One-way ANOVA: $p=0.90$, $F=0.19864$.

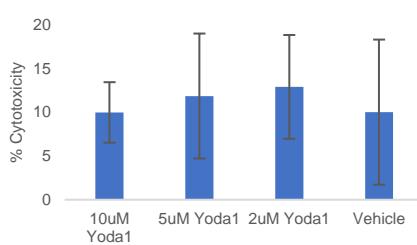
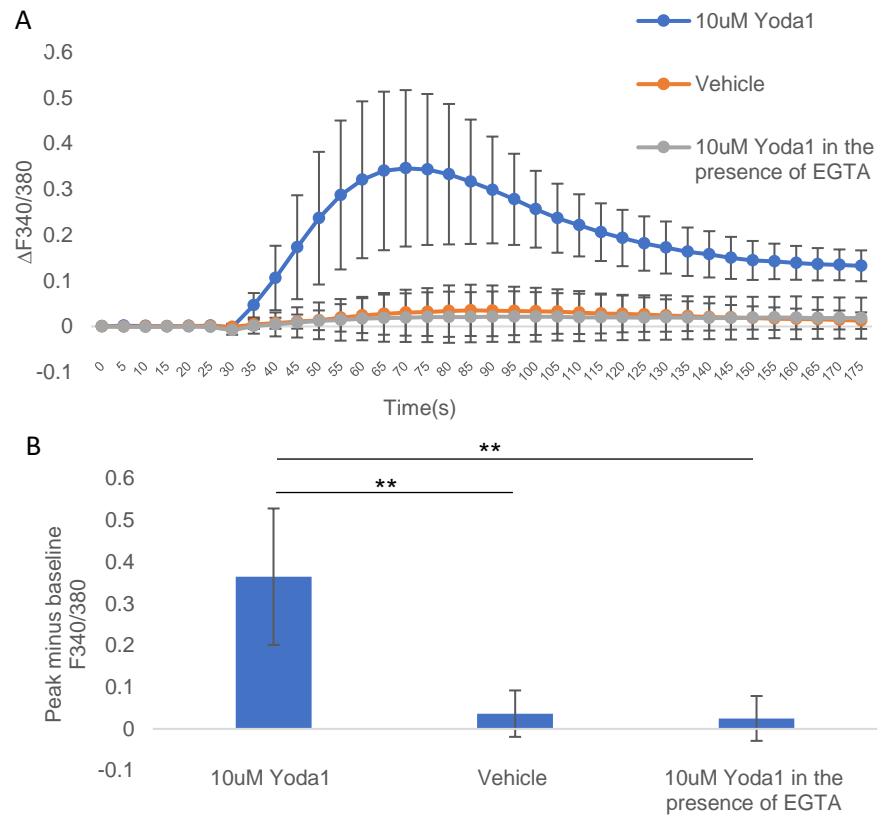


Figure 2. (A) Changes in intracellular Ca²⁺concentration ($\Delta F340/380$ ratio) after Yoda1 agonist was injected at 30 sec. Means and standard deviations are shown ($n = 4$). (B) Max F340/380 ratio in treated cells minus average of pre-injection baseline readings. One-way ANOVA: $p=0.0019$, $F=13.648$; Tukey HSD test: **, $p < 0.01$.



References:

- Blythe, N. M et al. (2019). Mechanically activated Piezo1 channels of cardiac fibroblasts stimulate p38 mitogen-activated protein kinase activity and interleukin-6 secretion. *Journal of Biological Chemistry*, 294(46), 17395–17408. <https://doi.org/10.1074/jbc.RA119.009167>
- Chen, X., & Thibeault, S. L. (2009). Novel Isolation and Biochemical Characterization of Immortalized Fibroblasts for Tissue Engineering Vocal Fold Lamina Propria. *Tissue Engineering Part C: Methods*, 15(2), 201–212. <https://doi.org/10.1089/ten.tec.2008.0390>
- Coste, B et al. (2010). Piezo1 and Piezo2 Are Essential Components of Distinct Mechanically Activated Cation Channels. *Science*, 330(6000), 55–60. <https://doi.org/10.1126/science.1193270>
- Fang, X.-Z. et al. (2021). Structure, kinetic properties and biological function of mechanosensitive Piezo channels. *Cell & Bioscience*, 11(1), 13. <https://doi.org/10.1186/s13578-020-00522-z>
- Foote, A. G. et al. (2022). Piezo1-expressing vocal fold epithelia modulate remodeling via effects on self-renewal and cytokeratin differentiation. *Cellular and Molecular Life Sciences*, 79(12), 591. <https://doi.org/10.1007/s00018-022-04622-6>
- Foote, A. G. et al. (2019). Tissue specific human fibroblast differential expression based on RNA sequencing analysis. *BMC Genomics*, 20(1), 308. <https://doi.org/10.1186/s12864-019-5682-5>
- Gray, S. D. (2000). CELLULAR PHYSIOLOGY OF THE VOCAL FOLDS. *Otolaryngologic Clinics of North America*, 33(4), 679–697. [https://doi.org/10.1016/S0030-6665\(05\)70237-1](https://doi.org/10.1016/S0030-6665(05)70237-1)
- Ilkan, Z. et al. (2017). Evidence for shear-mediated Ca²⁺ entry through mechanosensitive cation channels in human platelets and a megakaryocytic cell line. *Journal of Biological Chemistry*, 292(22), 9204–9217. <https://doi.org/10.1074/jbc.M116.766196>
- Lukacs, V. et al. (2015). Impaired PIEZO1 function in patients with a novel autosomal recessive congenital lymphatic dysplasia. *Nature Communications*, 6(1), 8329. <https://doi.org/10.1038/ncomms9329>
- Lungova, V. et al. (2020). Drainage of amniotic fluid delays vocal fold separation and induces load-related vocal fold mucosa remodeling. *Developmental Biology*, 466(1–2), 47–58. <https://doi.org/10.1016/j.ydbio.2020.08.003>
- Syeda, R. et al. (2015). Chemical activation of the mechanotransduction channel Piezo1. *eLife*, 4, e07369. <https://doi.org/10.7554/eLife.07369>

PODIUM SESSION 5

An Update on Conversation Training Therapy (CTT)

Amanda I Gillespie, PhD, CCC-SLP¹, Jackie Gartner-Schmidt, PhD, CCC-SLP²

¹ Emory University School of Medicine, Atlanta, GA, USA

² Carlow University, Pittsburgh, PA, USA

Keywords: voice, therapy, conversation

Abstract

Introduction & rationale: Since its inception in 2016, Conversation Training Therapy (CTT) has emerged as an innovative approach to voice therapy, using patient-led conversations as the sole therapeutic stimulus.

Objectives: CTT has undergone rigorous evaluation to assess efficacy, patient adherence, suitability, and the impact of its core elements on treatment outcomes. This presentation aims to provide a comprehensive update on the effectiveness and efficiency of CTT, drawing upon a synthesis of recent studies.

Methods: This presentation will distill findings from six peer-reviewed studies published over the last eight years, covering a broad spectrum of topics including the motor learning principles underpinning CTT, its proven efficacy in prospective trials, criteria for patient selection through stimulability testing, the benefits of group therapy delivery, enhancements in adherence through mobile app-supported practice, and the role of cognitive effort in spontaneous speech practice and vocal learning.

Results: Developed by six voice specialized speech-language pathologists, CTT is grounded in fundamental motor learning principles. Prospective trials have revealed that CTT demonstrates superior outcomes in acoustic, aerodynamic, auditory-perceptual, and patient-perceived measures at the conclusion of treatment and at a three-month follow-up, compared to traditional therapy methods. The speed of improvement in CTT is notably faster in patients who can discern changes in vocal effort and sound quality at pre-treatment stimulability assessment than those without such discernment. Patients participating in CTT via group and telehealth sessions reported significant reductions in Voice Handicap Index scores and treatment generalization, comparable to those in person one-on-one therapy settings. Additionally, the use of a mobile app to support CTT practice doubles the frequency of practice sessions and enhances practice fidelity of practice compared to unassisted practice.

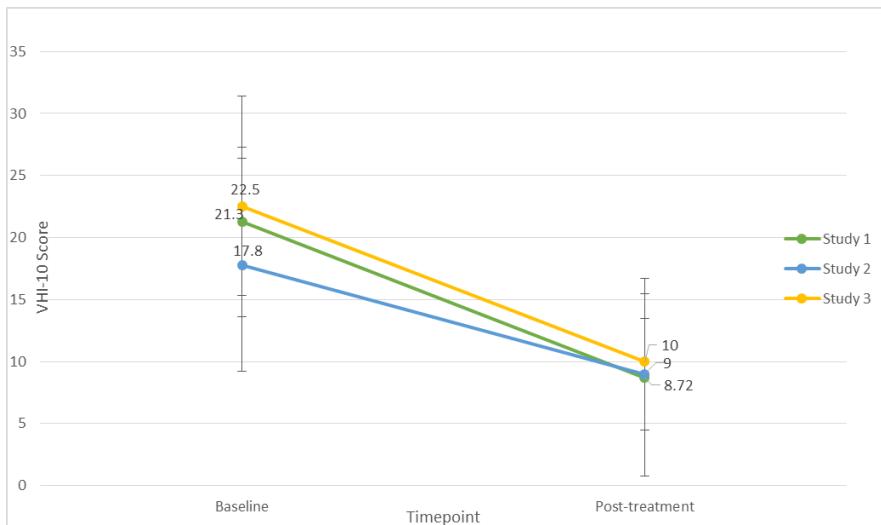


Figure 1 Change in VHI-10 Scores Pre-Post Treatment Across Multiple Studies using CTT. Study 1 = Gillespie et al., 2019; Study 2 = Gillespie, van Leer, Effects of mobile support on adherence to CTT, in preparation; Study 3 = Tripp, Gillespie, CTT delivered via virtual group therapy platform, in preparation

Conclusions: CTT is rapidly gaining recognition as a standard of care voice treatment approach across the United States. Its unique focus on facilitating the transfer of therapeutic concepts to spontaneous speech distinguishes it from other treatment methods, promising enhanced learning and sustained outcomes. Ongoing research continues to affirm its efficacy, effectiveness, and the significant impact of its treatment components on these results.

References:

1. Gartner-Schmidt, J., S. Gherson, E. R. Hapner, J. Muckala, D. Roth, S. Schneider and A. I. Gillespie (2016). The Development of Conversation Training Therapy: A Concept Paper. *J Voice* 30(5): 563-573.
2. Gartner-Schmidt, J. and A. I. Gillespie (2021). Conversation Training Therapy: Let's Talk It Through. *Semin Speech Lang* 42(1): 32-40.
3. Gillespie, A. I., J. Yabes, C. A. Rosen and J. L. Gartner-Schmidt (2019). Efficacy of Conversation Training Therapy for Patients With Benign Vocal Fold Lesions and Muscle Tension Dysphonia Compared to Historical Matched Control Patients. *J Speech Lang Hear Res*: 1-18.
4. Shelly S, Rothenberger SD, Gartner-Schmidt J, Gillespie AI. (2023) Assessing Candidacy for Conversation Training Therapy: The Role of Patient Perception. *J Voice*. Mar 10:S0892-1997(23)00044-9.
5. Gillespie AI, Shelly S, Xu H, Mayne GV. (2024) Patient Perception of Mental Effort in Voice Therapy. *J Voice*. Oct 31:S0892-1997(24)00246-7.
6. Iwarsson, J. (2015). Facilitating behavioral learning and habit change in voice therapy-theoretic premises and practical strategies. *Logoped Phoniatr Vocol*: 1-8.
7. Iwarsson, J., D. J. Morris and L. W. Balling (2017). Cognitive Load in Voice Therapy Carry-Over Exercises. *J Speech Lang Hear Res* 60(1): 1-12.
8. van Leer, E., E. R. Hapner and N. P. Connor (2008). Transtheoretical model of health behavior change applied to voice therapy. *J Voice* 22(6): 688-698.

The effect of listener feedback on sensorimotor adaptation of articulation and prosody

Kimberly L. Dahl¹, Manuel Díaz Cádiz¹, & Cara E. Stepp¹⁻³

¹ Department of Speech, Language, and Hearing Sciences, Boston University, Boston, MA, USA

² Department of Biomedical Engineering, Boston University, Boston, MA, USA

³ Department of Otolaryngology – Head and Neck Surgery, Boston University School of Medicine, Boston, MA, USA

Keywords: speech motor control, motor learning, intonation

Abstract

Introduction & rationale: Individuals rarely communicate without a listener, yet a primary method for studying speech motor control—altered auditory feedback—involves speaking in isolation. Speakers convey their message to the listener, in part, with effective articulation and prosody. Effective articulation requires accurate phoneme production, often defined by articulatory-acoustic features like the first vowel formant (F_1). Effective prosody involves fundamental frequency (f_0), amplitude, and rate—cues that can be traded to reach a prosodic target (Howell, 1993; Lieberman, 1960). Speakers subconsciously monitor their speech to ensure they have hit their intended targets. This self-monitoring allows the speaker to correct predictably recurring errors in F_1 or f_0 through sensorimotor adaptation, a key component of theories of speech motor control (Guenther & Vladusich, 2012; Houde & Nagarajan, 2011; Parrell et al., 2019).

Studies that use altered auditory feedback offer strong evidence of sensorimotor adaptation in both articulation and prosody, but thus far only in speech without communicative intent. Communicative intent is usually directed at a listener whose responses influence the speaker—including the speaker's choice of words and sentence structure (Brennan & Clark, 1996)—by indicating if the intended message was understood. Yet we do not know if listener feedback influences the distinct control mechanisms underlying articulation and prosody. The integrated nature of prosody may lead a speaker to respond to their own f_0 errors with less specificity than F_1 errors, as both the speaker and listener can draw on other prosodic cues to produce and perceive the intended message. Given the differences in control mechanisms and target specificity for articulation and prosody, the listener may exert unequal influences on these systems.

Objective: The purpose of this study was to understand how sensorimotor adaptation occurs in real-life communication by assessing the effect of listener feedback on articulatory and prosodic adaptation.

Methods: Thirty-six adults with no speech, language, or neurological disorders participated in an experimental task in which their auditory feedback and feedback from a listener were manipulated. Participants read aloud short sentences while seated before a computer monitor and wearing a microphone and earphones. During articulation tasks, F_1 of the first word was shifted downward 30%, resulting in a perceptual shift from /ɛ/ to /ɪ/ (e.g., “Belle” to “Bill”). During prosody tasks, participants were instructed to emphasize the first word of the sentence. The f_0 of the emphasized first word was shifted downward 200 cents, resulting in a perceptually unemphasized production. These acoustic shifts were applied to the middle 32 of 64 trials per condition and played through the participant's earphones in near-real time as they spoke. The intent of the shifts was to induce predictable errors in articulation and prosody and observe the degree to which the participant adapted to those errors.

In some conditions, participants read the sentences aloud to a listener. The task, as described to the participant, was for the listener to identify what the participant said in each trial when given the choice between the target sentence and another, similar-sounding sentence. That similar-sounding sentence differed only by the vowel (/ɛ/ or /ɪ/; articulation tasks) or stress (emphatic or neutral; prosody tasks) of the first word; that is, the listener's choices corresponded to the target sentence and the intended result of the applied acoustic shift. Participants were led to believe that listener feedback represented the real-time responses of a live listener; in reality, no such listener was present, and listener feedback was controlled by experimental scripts. The listener was incorporated with a pre-recorded video that mimicked an interaction by videoconference.

Listener feedback was presented in three conditions, including (1) an absent listener, as in a traditional altered auditory feedback design. When the listener was present, her feedback either (2) supported any adaptation that a participant exhibited under perturbed auditory feedback or (3) was indifferent to such adaptation. In other words, the listener responded in such a way that either confirmed that the participant had made or corrected an F_1 or f_0 error (condition 2) or that indicated the intended message was always understood regardless of any speech errors or corrections (condition 3).

Adaptive responses were quantified as F_1 or f_0 during shifted trials, when normalized to the 16 unshifted trials (i.e., baseline) that preceded the acoustic manipulation. The second formant (F_2) and amplitude were also extracted from the first word of each sentence to assess the specificity of adaptation. The effect of listener feedback on the magnitude, strength, and specificity of articulatory and prosodic adaptation was assessed via two mixed-effects and one repeated-measures analyses of variance.

Results: Participants adapted to the induced errors in both F_1 and f_0 by opposing the applied acoustic shift. Listener feedback had no statistically significant effect on the magnitude of prosodic adaptation but did affect articulatory adaptation; participants adapted less when the listener's understanding was indifferent to the F_1 errors or corrections. Articulatory adaptation was neither stronger nor more specific than prosodic adaptation, regardless of the presence of or type of listener feedback.

Conclusions: Speakers correct their errors in articulation and prosody, even if these errors do not undermine the speaker's success in communicating an intended message. However, in the articulatory domain, when communication is successful despite an error, the motor learning processes underlying error corrections (i.e., sensorimotor adaptation) may be weakened. This finding has implications for speech outcomes and speech interventions for people with impaired articulation. Despite differences in the acoustic targets, their flexibility, and speaker's responses to errors in the articulatory and prosodic domains (MacDonald et al., 2011; Patel et al., 2011; Villacorta et al., 2007), articulatory adaptation when communicating with a listener was neither stronger nor more specific than prosodic adaptation. The speaker's primary goal of conveying a message may therefore supersede the underlying differences in these speech domains.

References:

- Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(6), 1482–1493. <https://doi.org/10.1037/0278-7393.22.6.1482>
- Guenther, F. H., & Vladusich, T. (2012). A neural theory of speech acquisition and production. *Journal of Neurolinguistics*, 25(5), 408–422. <https://doi.org/10.1016/j.jneuroling.2009.08.006>
- Houde, J. F., & Nagarajan, S. (2011). Speech production as state feedback control. *Frontiers in Human Neuroscience*, 5, 1–14. <https://doi.org/doi.org/10.3389/fnhum.2011.00082>
- Howell, P. (1993). Cue trading in the production and perception of vowel stress. *The Journal of the Acoustical Society of America*, 94(4), 2063–2073. <https://doi.org/10.1121/1.407479>
- Lieberman, P. (1960). Some acoustic correlates of word stress in American English. *The Journal of the Acoustical Society of America*, 32(4), 451–454.
- MacDonald, E. N., Purcell, D. W., & Munhall, K. G. (2011). Probing the independence of formant control using altered auditory feedback. *The Journal of the Acoustical Society of America*, 129(2), 955–965. <https://doi.org/10.1121/1.3531932>
- 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. <https://doi.org/10.1371/journal.pcbi.1007321>
- Patel, R., Niziolek, C., Reilly, K., & Guenther, F. H. (2011). Prosodic adaptations to pitch perturbation in running speech. *Journal of Speech, Language, and Hearing Research*, 54(4), 1051–1059. [https://doi.org/10.1044/1092-4388\(2010/10-0162\)](https://doi.org/10.1044/1092-4388(2010/10-0162))
- Villacorta, V. M., Perkell, J. S., & Guenther, F. H. (2007). Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception. *The Journal of the Acoustical Society of America*, 122(4), 2306–2319. <https://doi.org/10.1121/1.2773966>

Effects of Delayed Auditory Feedback on Speech Kinematics

Christopher Dromey¹, Abbey Persons¹

¹ Communication Disorders, Brigham Young University, Provo, Utah, USA

Keywords: speech kinematics, articulation, delayed auditory feedback

Introduction & rationale:

Auditory feedback plays a critical role in speech production because it allows the nervous system to monitor and adjust motor behavior. This feedback enables self-correction and the maintenance of long-term speech quality. Delayed auditory feedback (DAF) manipulates an individual's perception of their own speech in real time. This technique has been widely used to investigate the link between perception and production. Research has shown that DAF can reduce stuttering during oral reading in controlled laboratory conditions. It has been suggested that DAF can improve fluency by disrupting stuttering patterns, facilitating smoother speech (Chon et al., 2021). DAF can also promote a reduction in speech rate in people with Parkinson's disease and increase intelligibility (Brendel, et al., 2004).

In this study, we sought to better understand the changes in articulatory movement that result from DAF. Although acoustic methods indirectly reflect vocal tract activity, electromagnetic articulography allows a more direct window into motor behavior.

Kinematic research has often focused on the extraction of 'point' measure like the distance moved and the speed of an articulator during a single phonetic gesture (e.g., Dromey & Ramig, 1998). An alternative is to use stroke measures to define speech movements, as outlined below in the Methods section. This approach allows automated segmentation of the articulatory movements in connected speech for any length of recording. Stroke measures provide metrics averaged over many sounds, reflecting the global control of articulatory movement, for example in response to changes in speech rate or vocal effort.

Objectives:

We set out to understand how speech movements change with different DAF latencies. The goal was to reveal global changes in speech motor control in typical speakers that may pave the way for future work with disordered populations where DAF might be applied clinically.

Methods:

The participants were 10 men and 10 women with typical speech, aged 18 to 29. Speech movements were captured with an electromagnetic articulograph. Along with a reference sensor on an eyeglass frame, sensors were positioned at midline on the upper lip (UL), lower lip (LL), at the tongue front (TF), and middle of the tongue (TM). A sensor was attached to the incisors to track the jaw (J). Kinematic data were sampled at 100 Hz and the audio signal at 22,050 Hz.

Speakers wore headphones to hear their speech with and without DAF. They read aloud a tongue twister four times under each DAF condition: "If a dog chews shoes, whose shoes does he choose?" Based on previous research (Brendel et al., 2004), we selected DAF latencies of 50 ms, 100 ms, and 150 ms, presented in randomly ordered blocks.

Stroke metrics were computed from speed plots in Matlab, which were based on the Euclidean distance between successive samples of X and Y points in the kinematic recording. Thus, the speed record reflected a change in position over time irrespective of direction. A stroke was defined as the movement between one speed minimum and the next throughout the speed history, as defined by Tasko and Westbury (2002).

Stroke measures were averages for the entire utterance and included the following: the stroke count (total number of articulatory strokes), along with means of the onset speed, peak speed, mean speed of the strokes, stroke distance, and stroke duration. Additionally, the 2-dimensional hull area in mm² was computed using the convex hull operation in Matlab. This area represented the boundary around all articulatory movements in the vertical and horizontal directions during the target sentence. These stroke metrics provide insight into articulator movement during continuous speech, presenting an overall or averaged view of speech movements across entire utterances, unlike traditional point measures that focus on single movements for specific syllables or phonemes.

A linear mixed model (LMM) analysis in SPSS was conducted to test for differences in the dependent variables across the conditions. DAF latency (0 ms, 50 ms, 100 ms, and 150 ms) was the fixed factor. Speaker sex was a fixed factor, and individual speaker was the random factor. Bonferroni-corrected pairwise multiple comparisons were used to test for significant differences between no DAF and the three latency conditions.

Results:

The stroke count did not change significantly across DAF conditions or differ between men and women for any of the articulators. There was a significant main effect of DAF condition on peak speed for the TF ($p < .001$), J ($p = .003$), and LL ($p < .001$). Pairwise comparisons revealed that the peak speed was lower for each DAF latency condition for the TF ($p < .001$). For the J, comparisons showed a decrease in peak speed for the 100 ms delay ($p = .035$) and the 150 ms delay ($p = .002$). For the LL, pairwise comparisons showed a significant decrease in peak speed for the 50 ms delay (p

= .034), for the 100 ms delay ($p = .027$) and the 150 ms delay ($p = .001$). For the J peak speed, there was a significant difference between men and women ($p = .031$), with a higher speed for women.

A significant main effect of DAF condition on TF stroke distance was found. TF stroke distance was significantly shorter for each DAF latency condition compared to no DAF. Specifically, TF distance decreased significantly for the 50 ms delay ($p = .009$), the 100 ms delay ($p = .019$), and the 150 ms delay ($p = .001$). For J stroke distance, there was a significant main effect of DAF condition, with pairwise comparisons showing that J distance was significantly shorter for the 150 ms delay compared to no DAF ($p = .030$). Similarly, LL distance was shorter for the 150 ms delay compared to no DAF ($p = .010$).

J distance was significantly shorter for men than women ($p = .014$). Additionally, LL distance was significantly shorter for men compared to women ($p = 0.019$). There was no significant effect of DAF on hull area for any of the articulators. However, hull area was significantly smaller for men compared to women for J ($p = .022$) and LL ($p = .019$).

Conclusions:

Previous work that included stroke measures for individuals speaking their first and second languages (Nissen et al., 2007) reported slower speeds in L2, suggesting increased effort and decreased articulatory automaticity. This is comparable with the findings of the current study on the effect of delayed auditory feedback (DAF), where peak speeds for the tongue-front (TF), jaw (J), and lower lip (LL) were significantly reduced under DAF conditions, suggesting a similar increase in effort and disruption in motor performance due to the external feedback manipulation. However, unlike the bilingual study, which reported slower stroke speeds without changes in stroke distance, the current findings revealed that stroke distances for TF, J, and LL decreased significantly under DAF, particularly at longer delay intervals (100 ms and 150 ms). This indicates that DAF not only affects speed but also the extent of articulatory movement.

The current study found significant sex differences in peak speeds, stroke distances, and hull areas for J and LL, with women generally showing higher speeds and longer stroke distances. These differences were consistent across both DAF and non-DAF conditions, suggesting inherent sex-based differences in speech kinematics for the tongue twister task. This finding emphasizes the importance of considering sex as a variable in speech production studies, which may provide further insights into the neuromuscular and articulatory characteristics across different populations.

The hull area, representing the spatial extent of articulatory movements in the x and y dimensions, could reasonably be presumed larger in men due to their generally larger anatomical structures. However, the findings of the current study contradict this assumption, showing that women had larger hull areas for both the jaw and lower lip compared to men. This counterintuitive finding could be attributed to women speaking more precisely, which involves greater and more controlled movements of the articulators. Women's speech precision may necessitate more extensive articulatory excursions, resulting in larger hull areas despite their relatively smaller anatomical structures. This precision in speech could be linked to sociolinguistic factors, where women are often observed to articulate more clearly and carefully. A study by Taylor et al. (2020) reported that women had higher spectral mean and kurtosis values for /s/ than men. While neither Taylor et al. (2020) nor the present study included perceptual ratings of speech precision, the data from both studies point to articulatory differences that appear to reflect greater precision in female speech.

The findings suggest that DAF disrupts motor performance and increases articulatory effort in structured and challenging tasks like tongue twisters. Future research could explore the perceptual impact of DAF-induced changes in a variety of speech contexts and further investigate the relationship between articulatory precision, effort, and listener perceptions of fluency.

References:

- Brendel, B., Lowit, A., & Howell, P. (2004). The effects of delayed and frequency shifted feedback on speakers with Parkinson's Disease. *Journal of Medical Speech-Language Pathology*, 12(4), 131–138.
- Chon, H., Jackson, E. S., Kraft, S. J., Ambrose, N. G., & Loucks, T. M. (2021). Deficit or difference? Effects of altered auditory feedback on speech fluency and kinematic variability in adults who stutter. *Journal of Speech, Language, and Hearing Research*, 64(7), 2539–2556.
- Dromey, C. & Ramig, L.O. (1998). Intentional changes in sound pressure level and rate: Their impact on measures of respiration, phonation and articulation. *Journal of Speech, Language, and Hearing Research*, 41, 1003-1018.
- Nissen, S. L., Dromey, C., & Wheeler, C. (2007). First and second language tongue movements in Spanish and Korean bilingual speakers. *Phonetica*, 64(4), 201–216.
- Tasko, S. M., & Westbury, J. R. (2002). Defining and measuring speech movement events. *Journal of Speech, Language, and Hearing Research*, 45(1), 127–142.
- Taylor, S., Dromey, C., Nissen, S. L., Tanner, K., Eggett, D., & Corbin-Lewis, K. (2020). Age-related changes in speech and voice: Spectral and cepstral measures. *Journal of Speech, Language, and Hearing Research*, 63(3), 647–660.

Sibilant differentiation following tongue cancer surgery: evidence from longitudinal articulatory-kinematics 6- and 12-months post-surgery

Thomas B. Tienkamp^{1,2,3}, Teja Rebernik⁴, Raoul Buurke^{1,3}, Nikki Hoekzema¹, Dan Mu^{1,3}, Katharina Polsterer^{1,3}, Hedwig Sekeres¹, Rob J.J.H. van Son^{5,6}, Martijn Wieling^{1,3}, Max J.H. Witjes², Sebastiaan A.H.J. de Visscher², and Defne Abur^{1,3}

¹Center for Language and Cognition Groningen, University of Groningen, Groningen, the Netherlands

²Department of Oral and Maxillofacial Surgery, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands

³Research School of Behavioral and Cognitive Neurosciences, University of Groningen, Groningen, the Netherlands

⁴Laboratoire de Phonétique et Phonologie, CNRS/Sorbonne Nouvelle, Paris, France

⁵Amsterdam Center for Language and Communication, University of Amsterdam, Amsterdam, the Netherlands

⁶Netherlands Cancer Institute, Amsterdam, the Netherlands

Keywords: Oral cancer; articulatory-kinematics; sibilants

Abstract

Introduction & rationale: Surgical treatment for tongue squamous cell carcinoma (TSCC) can negatively affect tongue mobility and control, leading to problems with swallowing and speech (Kreeft et al., 2009; Tienkamp et al., 2025). Surgical treatment for TSCC often distorts sibilants, such as /s/ and /ʃ/, reducing their acoustic and kinematic contrast. That is, the constriction location of /s/ shifts posteriorly, becoming more like /ʃ/ (Zhou et al., 2013). Yet, it is not clear how the time since surgical intervention for TSCC impacts these sibilant distortions, since speakers may develop compensatory strategies over time.

Objectives: The objective of this prospective case-control study was to assess articulatory-kinematic differentiation between the sibilants /s/ and /ʃ/ over time using electromagnetic articulography. Specifically, we assessed sibilant differentiation through a comparison of the Euclidean distance (ED), and anteroposterior (i.e., frontedness; AP) and superior-inferior (i.e., height; SI) contrast between the maximum constriction location of the tongue tip sensor for /s/ and /ʃ/. A more nuanced understanding of how surgery for TSCC affects the articulatory-kinematics of the tongue may facilitate the development of more effective phoneme-specific speech rehabilitation strategies. We predicted that post-surgery, the kinematic contrast between /s/ and /ʃ/, as measured by the Euclidean distance, AP- and SI-contrast, would become smaller as compared to pre-surgery and typical speakers.

Methods: Articulatory-kinematic data was collected from eleven native Dutch speakers pre-surgery and approximately 6- and 12-months post-surgery for TSCC (7 males; 4 females; mean age = 62 years old, age range = 40-77 years). Speakers were treated for a tumour located on the anterior two-thirds of the tongue ($n = 9$) or posterior one-third of the tongue ($n = 2$; OC02 and OC04). Two speakers (OC01 and OC02) received flap reconstruction; others were locally closed. For tumour size, which can range from T1 (smallest) to T4 (largest), individuals received treatment for T1 tumours ($n = 7$), T2 tumours ($n = 3$), T3 tumours ($n = 1$), or a carcinoma in situ ($n = 1$). Eleven typical speakers that were sex- and age-matched to those treated for TSCC participated as well (7 males; 4 females; mean age = 62 years; age range = 39-77 years). All speakers provided written informed consent and the study was approved by the institution's Medical Ethics Review Board.

Articulatory-kinematic data from five repetitions of the Dutch words 'sok' (sock, /sok/) and 'shock' (shock, /ʃok/) embedded in a carrier phrase were recorded with the NDI-Vox electromagnetic articulograph at 400 Hz (Rebernik et al., 2021). Data was collected using a sensor located approximately 1 cm from the anatomical tongue tip. Articulatory gestures were identified based on the velocity profile of the movement toward and away from the target constriction. Gesture trajectories of /s/ and /ʃ/ were normalised (z-scored) on a by-speaker basis to remove anatomical differences. From these normalised trajectories, we extracted the place of maximum constriction of both /s/ and /ʃ/ and computed the ED between the two target gestures in two-dimensional space. We also computed the contrast between the target gestures in the AP-dimension and SI-dimension in order to assess how each dimension contributes to potential changes in the ED. Data was analysed using a separate linear mixed-effects model for each outcome measure (ED, AP-contrast, and SI-contrast) with the outcome measure as a function of subgroup (typical, pre-surgery, 6-months post-surgery, and 12-months post-surgery) and by-speaker random intercepts.

Results: Figure 1 provides descriptive data of the contrast between /s/ and /ʃ/ in terms of the ED, AP-contrast, and SI-contrast. At 6-months post-surgery, individuals treated for TSCC had a smaller ED between /s/ and /ʃ/ compared to pre-surgery ($\beta = .37$, $p < .001$) and typical speakers ($\beta = .67$, $p = .001$). The ED between /s/ and /ʃ/ at 12-months post-surgery was significantly higher compared to 6-months post-surgery ($\beta = .27$, $p < .001$).

Dimension-specific analyses revealed that at 6-months post-surgery, the AP-contrast between /s/ and /ʃ/ was significantly smaller compared to pre-surgery ($\beta = .16, p = .006$) and typical speakers ($\beta = .79, p < .001$). No significant increase in AP-contrast was observed at 12-months post-surgery compared to 6-months post-surgery ($\beta = .01, p = .88$). In the SI-dimension, while there was a significant reduction 6-months post-surgery compared to pre-surgery ($\beta = .27, p < .001$), there was no significant difference 6-months post-surgery compared to typical speakers ($\beta = .07, p = .63$). At 12-months post-surgery, speakers showed a significantly larger IS-contrast between /s/ and /ʃ/ compared to 6-months post-surgery ($\beta = .3, p < .001$).

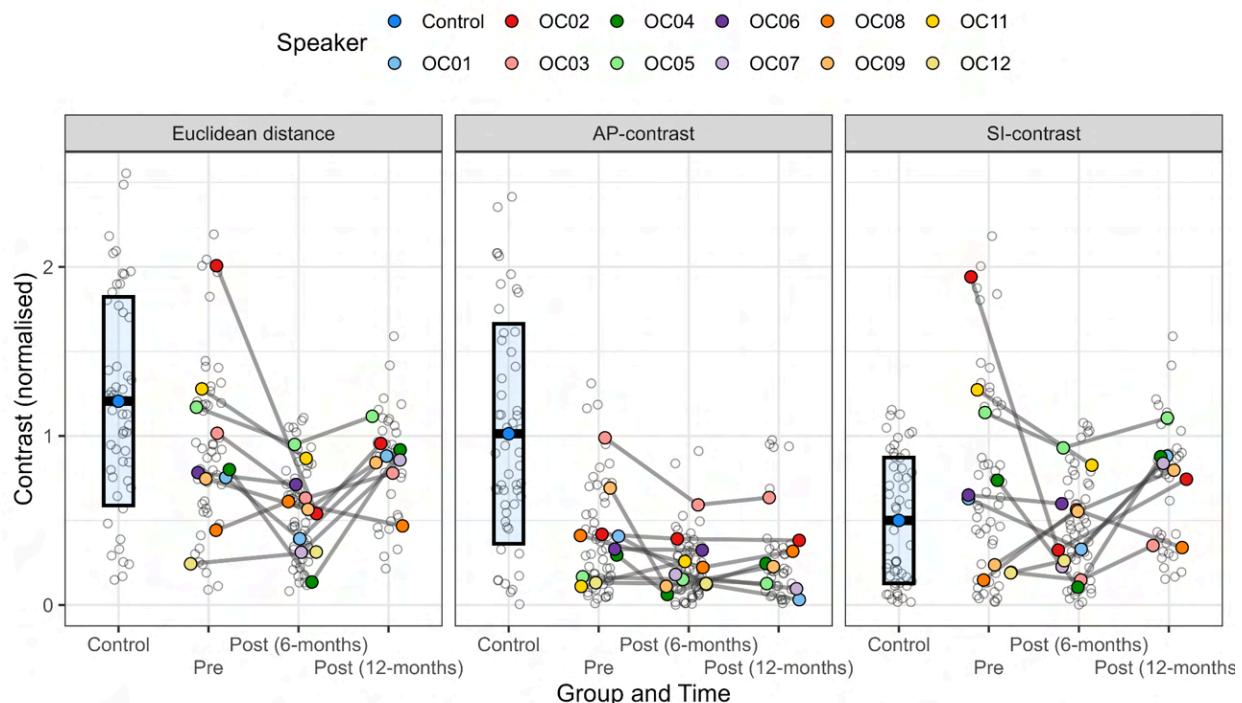


Figure 1. Contrast between the normalised constriction location of the tongue tip between /s/ and /ʃ/ in terms of the Euclidean distance, anterior-posterior contrast (AP-contrast), and superior-inferior contrast (SI-contrast). Each coloured data point represents the mean of the control group or an individual speaker, whereas unfilled data points represent the raw data. The blue crossbar for the controls represents a one standard deviation range. Data from individual speakers are connected by a grey line to show the development over time.

Conclusions: The current work assessed sibilant differentiation in typical speakers and individuals pre- and post-surgery for TSCC via electromagnetic articulography. In line with our prediction, our results indicated that, compared to typical speakers and pre-surgery, the sibilant contrast was reduced as a result of surgical treatment. While the reduced contrast was a result of both reductions in the AP and SI-dimensions, speakers increased their SI-contrast significantly at 12-months post-surgery. No increased AP-contrast was observed at 12-months post-surgery, signalling long-term challenges with tongue fronting. Taken together, our results showed both long-term reductions in tongue fronting as well as improvements in tongue height at 12-months post-surgery in sibilant differentiation in speakers surgically treated for TSCC. These results highlight the long-term adaptive capabilities of individuals, and suggest that studies need a follow-up time of at least 12-months to fully characterise post-surgery speech outcomes and compensatory strategies. We aim to supplement our current analysis with the kinematics of the jaw and lips in order to further explore compensatory movements that may arise after surgery for TSCC.

References:

- Kreeft, A. M., van der Molen, L., Hilgers, F. J., & Balm, A. J. (2009). Speech and swallowing after surgical treatment of advanced oral and oropharyngeal carcinoma: A systematic review of the literature. *European Archives of Oto-Rhino-Laryngology*, 266(11), 1687–1698. <https://doi.org/10.1007/s00405-009-1089-2>
- Rebernik, T., Jacobi, J., Tiede, M., & Wieling, M. (2021). Accuracy Assessment of Two Electromagnetic Articulographs: Northern Digital Inc. WAVE and Northern Digital Inc. VOX. *Journal of Speech, Language, and Hearing Research*, 64(7), 2637–2667. https://doi.org/10.1044/2021_JSLHR-20-00394
- Tienkamp, T. B., Rebernik, T., D'Cruz, R. A., Van Son, R. J. J. H., Wieling, M., Witjes, M. J. H., De Visscher, S. A. H. J., & Abur, D. (2025). Articulatory-kinematic changes in speech following surgical treatment for oral or oropharyngeal cancer: A systematic review. *International Journal of Language & Communication Disorders*, 60(1), e13148. <https://doi.org/10.1111/1460-6984.13148>
- Zhou, X., Woo, J., Stone, M., & Espy-Wilson, C. (2013). A cine MRI-based study of sibilant fricatives production in post-glossectomy speakers. *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 7780–7784. <https://doi.org/10.1109/ICASSP.2013.6639178>

PODIUM SESSION 6

Spatiotemporal indices of relative fundamental frequency for quantifying intraspeaker variability in vocal hyperfunction

Jenny Vojtech¹, Laura E. Toles², Daniel P. Buckley^{1,3}, Cara E. Stepp^{1,3,4}

¹ Department of Speech, Language, & Hearing Sciences, Boston University, Boston, MA 02215, USA

² Department of Otolaryngology – Head & Neck Surgery, UT Southwestern Medical Center, Dallas, TX 75390, USA

³ Department of Otolaryngology – Head & Neck Surgery, BU School of Medicine, Boston, MA 02118, USA

⁴ Department of Biomedical Engineering, Boston University, Boston, MA 02215, USA

Keywords: spatiotemporal index, relative fundamental frequency, vocal hyperfunction

Abstract

Introduction & Rationale: Vocal hyperfunction (VH) is a prevalent voice condition that affects approximately half of all cases referred to multidisciplinary voice clinics (Dworkin-Valenti et al., 2018; Roy, 2003). Clinical and research theories that excessive laryngeal muscle activity is the primary factor underlying dysphonic voice quality and persistent complaints of vocal effort dominate (e.g., Hanschmann et al., 2010; van Mersbergen et al., 2021). Various factors—such as improper voice use, psychological predisposition, and laryngeal biomechanics—are thought to lead to hyperfunctional vocal behaviors. These behaviors are characterized as occurring independently or as a response to vocal injury (Hillman et al., 2020). When VH occurs without injury, it is called "non-phonotraumatic" VH (**NPVH**) and is often diagnosed as muscle tension dysphonia. In the presence of vocal fold lesions such as polyps or nodules, VH is classified as "phonotraumatic" (**PVH**). Clinically assessing VH is challenging due to the subjective nature of current evaluation methods and the inherent heterogeneity of the condition. Relative fundamental frequency (**RFF**) has shown promise as an objective, acoustic tool to supplement voice assessments but has not been widely adopted in clinic, partly due to intraspeaker variability in mean values (Roy et al., 2016). Although has traditionally been viewed as a limitation, few studies have explored its underlying causes or potential solutions. *But what if this variability provides meaningful insights into the nature of VH?*

Increased intraspeaker variability in voice acoustics (Belsky et al., 2020), aerodynamics (Belsky et al., 2020; Gillespie et al., 2013; Higgins et al., 1999), and articulatory kinematics (Shipurkar et al., 2023) has been observed in speakers with VH relative to those without VH. Research similarly shows greater intraspeaker variability in speech production relative to those without VH (McKenna et al., 2020), suggesting inaccuracies in feedforward sensorimotor control of speech. PVH has been further associated with even more variable and inefficient speech production due to vocal fold lesions that are thought to contribute to an inherently unstable physical system (Crocker et al., 2024; Free et al., 2023). This instability may increase reliance on feedback control mechanisms, leading to greater speech variability in people with PVH compared to those with NPVH. Rather than treating this variability as an obstacle to accurate measurement, it is possible that intraspeaker variability may reflect important characteristics of VH rather than impediments to accurate assessment. As mean RFF values are thought to reflect the level of laryngeal tension during phonation, intraspeaker RFF variability may indicate deficiencies in laryngeal motor control during phonation. Thus, we propose that this variability may be useful for improving the discriminative power of RFF in VH assessment.

Objectives: This study sought to explore whether measures of RFF variability could improve the diagnostic utility of RFF in identifying and discriminating VH subtypes. Two distinct measures of intraspeaker variability were investigated: a basic measure, standard deviation (SD), and a more complex measure, the spatiotemporal index (STI), which assesses movement pattern stability over repeated performances of the same motor task (e.g., speech production).

Methods: Adults with PVH ($n = 44$), NPVH ($n = 44$), and age- and sex-matched controls ($n = 44$) participated in the study, for a total of 132 participants. Participants were matched within a 5-year age range across groups. Individuals with VH required a diagnosis consistent with PVH (structural pathology such as nodules or polyp with observed VH) or NPVH (muscle tension dysphonia) by a referring laryngologist. Speech samples were collected from all participants as they produced repetitions of three vowel–voiceless consonant–vowel tokens (/afa/, /ifi/, /ufu/) with equal stress on each vowel at their typical pitch and loudness. Acoustic signals were processed using an open-source algorithm called *aRFF-AP* (Vojtech et al., 2019) to generate RFF traces for each token. RFF traces were averaged at voice offset (cycle 10) and voice onset (cycle 1) to produce mean RFF values ($M_{\text{offset}10}$, $M_{\text{onset}1}$). Two measures of RFF variability—SD and STI—were calculated alongside mean RFF values. SD was calculated across RFF traces at voice offset cycle 10 ($SD_{\text{offset}10}$) and onset cycle 1 ($SD_{\text{onset}1}$). STI was calculated for voice offset and onset using methodology from Smith et al. (2000)—with a gamma-based correction applied as in Wisler et al. (2022) to address potential bias from the number of traces—resulting in two values per participant: one for offset (STI_{offset}) and one for onset (STI_{onset}).

A permutational multivariate analysis of variance (**PERMANOVA**) was conducted to examine how the three groups (control, PVH, NPVH) differed on each of the six RFF measures ($M_{\text{offset}10}$, $M_{\text{onset}1}$, $SD_{\text{offset}10}$, $SD_{\text{onset}1}$, STI_{offset} , STI_{onset}). Significant measures were entered along with demographic parameters (age, sex) into hierarchical multinomial logistic regression models using a training set ($n = 102$). Final model equations were then applied to an

independent test set ($n = 30$) to predict group membership. One-vs-rest receiver operating characteristic (ROC) curve analyses were performed to evaluate discriminative performance on the test dataset. Final metrics of area under the curve (AUC), sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated for each group to evaluate model performance.

Results: Mean ($M_{\text{offset}10}$, $M_{\text{onset}1}$) and STI (STI_{offset} , STI_{onset}) measures showed significant group differences, whereas SD did not ($p \geq .05$). The PVH group showed significantly lower $M_{\text{offset}10}$ values than the control and NPVH groups. Both VH groups had significantly lower $M_{\text{onset}1}$ values than the control group. STI_{offset} was significantly higher in the NPVH group than in the PVH group and STI_{onset} was significantly lower in controls compared to both VH groups.

A single hierarchical multinomial logistic regression model was constructed, incorporating the mean and STI measures that showed significant group effects. Block 1 of the model—which included basic demographic parameters of age and sex—did not improve model fit for discerning group compared to the null model with no predictors ($\chi^2_{\text{change}} = 3.57$, $p = .167$, Pseudo- $R^2_{\text{adj}} = .04$). Mean RFF values entered in Block 2 led to statistically significant improvements in model fit ($\chi^2_{\text{change}} = 27.63$, $p < .001$) with Pseudo- $R^2_{\text{adj}} = .59$. The introduction of STI values into the model in Block 3 also resulted in statistically significant improvements in fit ($\chi^2_{\text{change}} = 8.91$, $p = .012$) with Pseudo- $R^2_{\text{adj}} = .68$.

One-vs-rest ROC curves were generated using the predictions from each test set (Fig. 1), yielding a median AUC of 0.78. For NPVH, the model achieved an AUC of 0.78 (sensitivity = 0.70, specificity = 0.75), consistent with an acceptable classifier. Classification performance was lowest for PVH, with an AUC of 0.69 (sensitivity = 0.60, specificity = 0.75). Despite lower PPVs of 0.58 for NPVH and 0.55 for PVH, the consistently high NPV of 0.75 across VH groups suggests the model is more effective at ruling out VH than confirming it. For controls, the AUC was slightly higher at 0.86 (sensitivity = 0.70, specificity = 0.99), consistent with a strong classifier. The PPV and NPV of the control group were both 0.99, indicating excellent reliability in ruling out control cases.

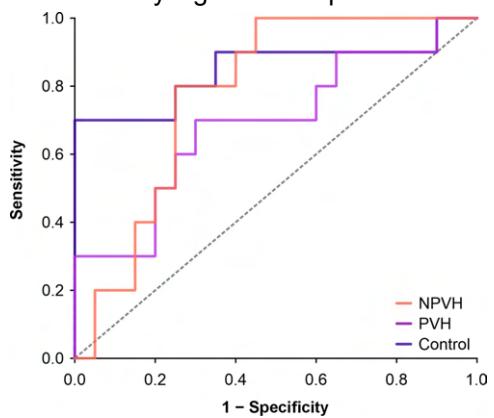


Figure 1. Model ROC performance on the test set.

Conclusions: This study demonstrates that incorporating RFF variability in both time and magnitude—via STI—provides valuable, statistically significant information about the presence and type of VH, offering insights beyond what mean values yield alone. These results underscore the clinical value of considering RFF variability in the characterization of hyperfunctional voice disorders which may, in turn, lead to more nuanced evaluations of VH toward improving patient outcomes. Future research should focus on further validating these metrics across larger and more diverse populations, as well as investigating their sensitivity to changes in vocal function over time and in response to treatment.

References

- Belsky, M. A., et al. (2020). Do phonatory aerodynamic and acoustic measures in connected speech differ between vocally healthy adults and patients diagnosed with muscle tension dysphonia? *Journal of Voice*, 35(4), 663.e1–663.e7.
- Crocker, C., et al. (2024). Relationships Between Vocal Fold Adduction Patterns, Vocal Acoustic Quality, and Vocal Effort in Individuals with and without Hyperfunctional Voice Disorders. *Journal of Voice*, S0892-1997(23)00405-8.
- Dworkin-Valenti, J. P., et al. (2018). Pathophysiologic perspectives on muscle tension dysphonia. *Archives of Otolaryngology and Rhinology*, 4(1), 1–10.
- Free, N., Stemple, J. C., Smith, J. A., & Phyland, D. J. (2023). Variability in Voice Characteristics of Female Speakers with Phonotraumatic Vocal Fold Lesions. *Journal of Voice*, S0892-1997(23)00017-6.
- Gillespie, A. I., Gartner-Schmidt, J., Rubinstein, E. N., & Abbott, K. V. (2013). Aerodynamic Profiles of Women with Muscle Tension Dysphonia/Aphonia. *Journal of Speech, Language, and Hearing Research*, 56(2), 481–488.
- Hanschmann, H., Lohmann, A., & Berger, R. (2010). Comparison of Subjective Assessment of Voice Disorders and Objective Voice Measurement. *Folia Phoniatrica et Logopaedica*, 63(2), 83–87.
- Higgins, M. B., Chait, D. H., & Schulte, L. (1999). Phonatory air flow characteristics of adductor spasmodic dysphonia and muscle tension dysphonia. *Journal of Speech, Language, & Hearing Research*, 42(1), 101–111.
- 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.
- McKenna, V. S., et al. (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.
- Roy, N. (2003). Functional dysphonia. *Current Opinion in Otolaryngology & Head and Neck Surgery*, 11(3), 144–148.
- Roy, N., Fetrow, R.A., Merrill, R.M., & Dromey, C. (2016). Exploring the clinical utility of relative fundamental frequency as an objective measure of vocal hyperfunction. *Journal of Speech, Language, & Hearing Research*, 59(5), 1002–1017.
- Shipurkar, R., White, M., & Katz, W. (2023). Normalization of speech kinematic data for characterizing primary muscle tension dysphonia. *Acoustics 2023*, Sydney, Australia.
- van Mersbergen, M. R., Beckham, B. H., & Hunter, E. J. (2021). Do We Need a Measure of Vocal Effort? Clinician's Report of Vocal Effort in Voice Patients. *Perspectives of the ASHA Special Interest Groups*, 6(1), 69–79.
- Vojtech, J. M., et al. (2019). Refining algorithmic estimation of relative fundamental frequency: Accounting for sample characteristics and fundamental frequency estimation method. *J. Acoust. Soc. Am.*, 146(5), 3184–3202.

Differences in ambulatory sympathetic arousal between patients with vocal hyperfunction and vocally healthy controls

Jeremy Wolfberg^{1,2}, Jameson C. Cooper¹, Daryush Mehta^{1,2,3}, Nelson Roy⁴, Robert E. Hillman^{1,2,3}, Jarrad H. Van Stan^{1,2,3}

¹ Center for Laryngeal Surgery and Voice Rehabilitation, Massachusetts General Hospital, Boston, MA, USA

² School of Health and Rehabilitation Sciences, MGH Institute of Health Professions, Boston, MA, USA

³ Harvard Medical School, Harvard University, Boston, MA, USA

⁴ Department of Communication Sciences and Disorders, The University of Utah, Salt Lake City, UT, USA

Keywords: Sympathetic Arousal, Electrodermal Activity, Ambulatory Monitoring

Abstract

Introduction & rationale:

Vocal hyperfunction (VH) consists of imbalanced and/or elevated musculoskeletal activity (Hillman et al., 2020) and represents some of the most common voice disorders (Bhattacharya, 2014). VH can be divided into two subcategories: phonotraumatic vocal hyperfunction (PVH), characterized by the presence of trauma on the vocal fold(s), and non-phonotraumatic vocal hyperfunction (NPVH), characterized by dysphonia in the absence of any physical or neurological phonation-related impairment (Hillman et al., 2020). Elevated sympathetic arousal, the neural system underlying the “fight or flight” response, is a proposed mechanism associated with hyperfunctional voice disorders. For example, the Trait Theory of Voice Disorders hypothesizes that heightened sympathetic arousal occurs in both patients with PVH and patients with NPVH, albeit via different neural substrates called the Behavioral Activation System and Behavioral Inhibition System, respectively (Roy & Bless, 2000). Previous studies showed that patients with VH self-reported more sympathetic nervous system symptoms than vocally healthy controls (Demmink-Geertman & Dejonckere, 2002), but no studies to the authors’ knowledge have objectively measured differences in sympathetic arousal in daily life. Electrodermal activity (EDA) measures overall sympathetic arousal by tracking sweat gland activity and can be collected in an ambulatory setting with smartwatch technology (Bouscain, 2012).

Objectives:

The primary aim of this study is to investigate differences in ambulatory measures of sympathetic nervous system activity using EDA among patients diagnosed with NPVH, PVH, and vocally healthy controls.

Methods:

121 participants consented to participate in this study, including 57 patients with PVH, 30 patients with NPVH, and 34 endoscopically confirmed vocally healthy controls. All patients and controls completed ambulatory EDA monitoring prior to receiving any behavioral or surgical treatment for their voice. Participants wore the smartwatch device for an average (standard deviation [SD]) of 5.56 (1.67) days. EDA measures included skin conductance response (SCR) count per minute, SCR amplitude mode, and skin conductance level (SCL) mode. Independent sample t-tests assessed significant differences between NPVH versus controls, PVH versus controls, and PVH versus NPVH. For significant findings, Cohen’s *d* effect sizes estimated the magnitude of difference.

Results:

Patients with NPVH (Mean [M] = 6.00 microsiemens [μ Si], SD = 10.97) and patients with PVH (M = 4.17 μ Si, SD = 6.60) exhibited significantly higher SCL mode compared to the vocally healthy controls (M = 2.43 μ Si, SD = 2.55). Patients with NPVH exhibited a small-to-medium effect size; $t(32) = 1.74$, $p = 0.045$, $d = 0.46$. Patients with PVH exhibited a small effect size; $t(79) = 1.77$, $p = 0.040$, $d = 0.32$. There were no statistically significant differences between the patient groups. There were also no statistically significant differences between the groups for SCR count per minute and SCR amplitude mode.

Conclusions:

Due to the phasic and transient nature of SCR, longer-term tonic measures of SCL may represent a more robust method for estimating sympathetic activation in an ambulatory setting. To our knowledge, this is the first study empirically supporting that patients with VH exhibit elevated sympathetic arousal in daily life. These results potentially indicate that patients with VH spend more time in the “fight or flight” response. Higher sympathetic activation has also been associated with chronic hypervigilance and/or stress (Brosschot, 2010; Thayer & Lane, 2009), which aligns with literature demonstrating relationships between VH and somatic amplification (Roy & Bless, 2000; Shembel et al., 2024), as well as laryngeal responses to stress (Abur et al., 2023; Helou et al., 2020). Due to the heterogenous nature of NPVH and PVH, future work could also use EDA data to aid in the phenotyping of patients in these diagnostic groups, potentially aiding the development of precision medicine.

References:

- Abur, D., MacPherson, M. K., Shembel, A. C., & Stepp, C. E. (2023). Acoustic measures of voice and physiologic measures of autonomic arousal during speech as a function of cognitive load in older adults. *Journal of Voice*, 37(2), 194–202. <https://doi.org/10.1016/j.jvoice.2020.12.027>
- Bhattacharyya, N. (2014). The prevalence of voice problems among adults in the United States. *The Laryngoscope*, 124(10), 2359–2362. <https://doi.org/10.1002/lary.24740>
- Bouscain, W. (2012). *Electrodermal activity* (2nd edition). Springer.
- Brosschot, J. F. (2010). Markers of chronic stress: Prolonged physiological activation and (un)conscious perseverative cognition. *Psychophysiological Biomarkers of Health*, 35(1), 46–50. <https://doi.org/10.1016/j.neubiorev.2010.01.004>
- Demmink-Geertman, L., & Dejonckere, P. H. (2002). Nonorganic habitual dysphonia and autonomic dysfunction. *Journal of Voice*, 16(4), 549–559. [https://doi.org/10.1016/S0892-1997\(02\)00130-3](https://doi.org/10.1016/S0892-1997(02)00130-3)
- Helou, L. B., Jennings, J. R., Rosen, C. A., Wang, W., & Verdolini Abbott, K. (2020). Intrinsic laryngeal muscle response to a public speech preparation stressor: Personality and autonomic predictors. *Journal of Speech, Language, and Hearing Research*, 63(9), 2940–2951. https://doi.org/10.1044/2020_JSLHR-19-00402
- 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. https://doi.org/10.1044/2020_AJSLP-20-00104
- Roy, N., & Bless, D. M. (2000). Personality traits and psychological factors in voice pathology: A foundation for future research. *Journal of Speech, Language, and Hearing Research*, 43(3), 737–748. <https://doi.org/10.1044/jslhr.4303.737>
- Shembel, A. C., Mau, T., Zafereo, J., Morrison, R., Crocker, C., Moore, A., & Khan, A. (2024). Laryngeal and global somatosensation in primary muscle tension dysphonia. *Journal of Voice*. <https://doi.org/10.1016/j.jvoice.2024.08.003>
- Thayer, J. F., & Lane, R. D. (2009). Claude Bernard and the heart–brain connection: Further elaboration of a model of neurovisceral integration. *The Inevitable Link between Heart and Behavior: New Insights from Biomedical Research and Implications for Clinical Practice*, 33(2), 81–88. <https://doi.org/10.1016/j.neubiorev.2008.08.004>

EFFECTS OF COGNITIVE LOAD ON VOCAL MOTOR CONTROL: CHANGES IN SUBGLOTTAL PRESSURE AND LARYNGEAL MUSCLE ACTIVATION

Rocío B. Ortega^{1,2}, Jesús A. Parra², Nicolás F. Quinteros^{1,2}, Emiro Ibarra², Cara E. Stepp³, Matías Zañartu^{1,2}

¹ Department of Electronic Engineering, Universidad Técnica Federico Santa María, Valparaíso, Chile

² Advanced Center for Electrical and Electronic Engineering, Universidad Técnica Federico Santa María, Valparaíso, Chile

³ Department of Speech, Language, and Hearing Sciences and Department of Biomedical Engineering, Boston University, Boston, Massachusetts.

Keywords: Cognitive load, laryngeal motor control, feedback control model.

Abstract

Introduction & rationale:

Cognitive load (CL) refers to the mental effort required to learn or perform a specific task. Performing tasks under high CL provides valuable insights into an individual's neurological and physiological state. The growing interest in using voice as an indicator of CL arises from its potential to facilitate the early detection of impairments in intellectual, verbal, perceptual, executive, motor, and emotional functions (Ehrenberg et al., 2020). Previous studies have shown that increased laryngeal tension and sound pressure levels (SPL), along with changes in cepstral peak prominence and the low-to-high spectral energy ratio, are effects of elevated CL (Pyfrom et al., 2023). These findings suggest that CL may alter vocal motor control and affect voice quality. Therefore, estimating and analyzing motor inputs such as intrinsic laryngeal muscle activations of the cricothyroid (CT) and arytenoid (TA), and subglottal pressure (P_s) may provide deeper insight into how CL influences motor control mechanisms and manifests in voice production. In this context, we hypothesize that an increase in CL will lead to systematic changes in motor control mechanisms, which will, in turn, affect acoustic features such as fundamental frequency (f_0) and SPL. These changes can be quantified using a feedback control model that performs inverse mapping of acoustic outputs (f_0 and SPL) to motor control features (P_s , TA, and CT).

Objectives:

To determine the effect of cognitive load on voice production by estimating laryngeal motor inputs through a feedback vocal control model.

Methods:

A CL dataset of speech signals recorded from 31 female participants was utilized to analyze CL during experiments based on the Color-Word Stroop task. In this task, each participant read the same sentence aloud under two conditions: the Congruent Condition (CC), where the font color and the word meaning matched, representing normal CL, and the Incongruent Condition (IC), where the font color and the word meaning did not match, thereby increasing CL (Dahl & Stepp, 2023). To facilitate temporal signal comparison, recordings were segmented to include only phonated sections during CL perturbation.

Laryngeal motor features were estimated using a feedback control model that simulates vocal control through a two-step mapping process. First, motor inputs (CT, TA, and P_s) are mapped to acoustic outputs (f_0 and SPL) via a polynomial regressor trained on thousands of simulations of the triangular body-cover model (Sorolla et al., 2024). Then, an inverse mapping refines motor control by minimizing the error between predicted acoustic outputs and auditory targets, which correspond to the f_0 and SPL trajectories extracted from microphone signals using Praat (30 ms window). This error is transformed into corrective motor updates through the inverse Jacobian. Once the model's outputs align with the targets, the resulting motor input trajectories are assumed to represent actual laryngeal motor commands. Figure 1 shows the temporal evolution of SPL and f_0 for a single subject, comparing model predictions with auditory targets and the corresponding motor inputs (CT, TA, and P_s).

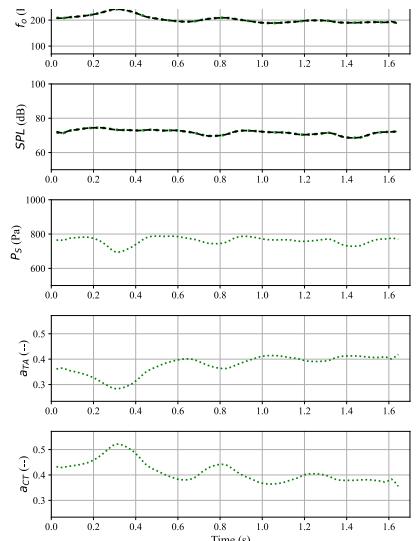


Figure 1: Temporal evolution of the input-output response of the feedback control model (green) adapting to match the acoustic target signals (black). From top to bottom, the acoustic features (F_0 and SPL) represent the model input, while the motor control parameters (P_s , TA, and CT) represent the model output.

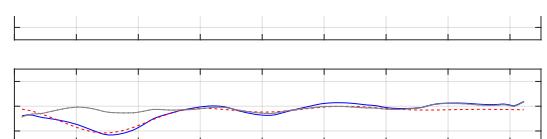


Figure 2: SSA decomposition of the motor feature. The blue line corresponds to the motor feature estimated by the feedback model, the red line shows the reconstructed trajectory based on components explaining 80% of the variance, and the gray line depicts the residual.

Singular Spectrum Analysis (SSA) was applied to each set of laryngeal motor features (CT, TA, P_s), as shown in Figure 2. This method decomposed the signals, allowing for the preservation of underlying trends and oscillations that tribute to the most relevant aspects of each feature's trajectory. For subsequent analysis, 80% of the variance explained by the components was used to reconstruct the signal, ensuring that the retained components adequately represented the differences between the CC and IC conditions.

Results:

In Figure 3, the mean trajectories of the estimated motor features show notable differences between the CC and IC conditions. In particular, a decrease in the activation of the TA muscle and an increase in the CT muscle activation and P_s, which also exhibit fluctuations over time, are observed under CL. Given the antagonistic relationship between the CT and TA muscles, these changes may reflect increased muscular tension to maintain vocal motor control stability under cognitive demand.

Furthermore, activation of the CT and TA muscles indicates that both conditions initially follow a similar trajectory, suggesting a possible delay in the laryngeal motor response under CL. This could be related to a delayed compensation mechanism in vocal control.

To further support these findings, an analysis of variance (ANOVA) was performed to assess statistical differences between the two groups, considering both acoustic and motor features. As shown in Table 1, significant effects ($p < 0.05$) were found for P_s, TA muscle activation, SPL and f_o, indicating that the differences observed in the mean motor trajectories are statistically significant. Although the acoustic (f_o and SPL) and motor (TA and P_s) features are correlated, the size effect (η_p^2) observed between conditions are primarily driven by P_s and TA activation. These results suggest that vocal production changes under CL are influenced by laryngeal motor control.

Conclusions:

This study investigates the influence of CL on voice production by analyzing laryngeal motor inputs through a feedback vocal control model. Results demonstrate significant alterations in motor features, such as increased P_s and CT muscle activation, although not statistically significant, and reduced TA muscle activation under CL, suggesting greater muscular tension to maintain vocal control stability. These findings highlight the crucial role of laryngeal motor control in vocal production under varying cognitive demands. Additionally, future research will evaluate the relevance of incorporating word duration and phoneme to study whether the delay mechanisms in CT and TA muscle is related to this factor.

Acknowledgements:

This research was supported by the National Institutes of Health (NIH) National Institute on Deafness and Other Communication Disorders grant P50 DC015446, ANID grants FONDECYT 1230828, BASAL AFB240002 and National Scholarship 22231668.

References:

- Dahl, K. L., & Stepp, C. E. (2023). Effects of Cognitive Stress on Voice Acoustics in Individuals With Hyperfunctional Voice Disorders. *American Journal of Speech-Language Pathology*, 32(1), 264–274. https://doi.org/10.1044/2022_AJSLP-22-00204.
- Ehrenberg, A. J., Khatun, A., Coomans, E., Betts, M. J., Capraro, F., et al. (2020). Relevance of biomarkers across different neurodegenerative diseases. *Alzheimer's Research & Therapy*, 12(1), 56. <https://doi.org/10.1186/s13195-020-00601-w>.
- Pyfrom, M., Lister, J., & Anand, S. (2023). Influence of Cognitiva Load on Voice Production: A Scoping Review. *Journal of Voice*, S0892199723002710. <https://doi.org/10.1016/j.jvoice.2023.08.024>.
- Sorolla, C., Parra, J. A., Ibarra, E., Alzamendi, G., & Zañartu, M. (2024). Forward mapping estimation for laryngeal motor control using machine learning techniques. *Proceedings of 13th International Conference on Voice Physiology and Biomechanics (ICVPB)*.

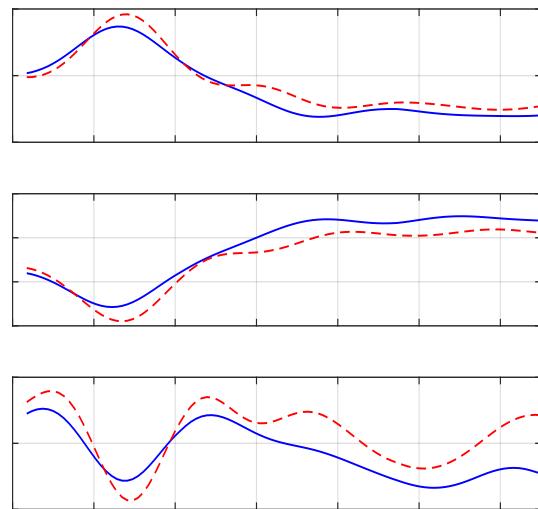


Figure 3: Mean trajectories of motor features for both CC (blue) and IC condition (red).

Feature	df	F	p	η_p^2
CT (-)	1	2.073	0.151	0.010
TA (-)	1	5.969	0.015	0.029
P _s (Pa)	1	27.931	<0.001	0.124
f _o (Hz)	1	4.840	0.029	0.024
SPL (dB)	1	26.872	<0.001	0.119

Table 1: Results of ANOVA test in acoustic and motor features for both CC and IC conditions.

PODIUM SESSION 7

Transverse Ultrasound Imaging of the Phonating Larynx

Mark K. Tiede¹, D.H. Whalen^{2,3}, Laura L. Koenig^{3, 4}

¹ Dept. of Psychiatry, Yale University, New Haven, CT, USA

² Speech-Language-Hearing Sciences, City University of New York, New York, NY, USA

³ Child Study Center, Yale University, New Haven, CT, USA

⁴ Communication Sciences & Disorders, Adelphi University, Garden City, NY, USA

Keywords: transverse ultrasound, devoicing gesture, laryngeal kinematics

Abstract

Introduction & rationale: Ultrasound imaging of speech articulation, primarily midsagittal tongue posture, has become well-established in recent years due to its ease of use and non-invasive qualities, but its application to laryngeal imaging to date has been limited. In general the larynx is observed using videoendoscopic methods, which are invasive, perturbative, and poorly tolerated by many speakers. In addition, the view from above they provide is difficult to quantify given the typically uncalibrated and variable height difference between the camera and the laryngeal structures. Accordingly ultrasound approaches for imaging the larynx can potentially provide a useful non-invasive alternative. Some authors have successfully measured changes in laryngeal height using coronal ultrasound (Matsunaga, Niimi, & Hirose, 1991; Poh & Moisik, 2019), and comparison of glottographic and ultrasound images supported an assessment of laryngeal height during production of Mandarin tones (Moisik, Lin, & Esling, 2014). Ultrasound imaging of the larynx has also been employed to study swallowing and vocal fold motility pre/post-surgery (Allen et al., 2021). However, ultrasound imaging of laryngeal structures during speech production has to our knowledge not been extensively reported. Pilot work reported here demonstrates that transverse ultrasound can track consonant devoicing gestures by imaging ab/adductory cycles of the ventricular (false) folds, known to be reliably correlated with movements of the true folds (Pressman & Kelemen, 1955; Poletto et al., 2004). In addition, portions of the arytenoid cartilages protruding into the ultrasound imaging plane are reliably visualized and can be used to track their coordination with supralaryngeal articulation.

Objectives: Our goal with this work is not to image the actual vibration of the vocal folds, but rather to better observe the complex laryngeal structures that control and support them during speech, filling a substantial gap in our knowledge of the mechanisms of speech production with both linguistic and clinical implications.

Methods: Ultrasound was used to visualize laryngeal devoicing gestures in nonsense utterances produced by native speakers of American English. The utterances had V₁CV₁ structure, with C drawn from [h], [?], [p], and [b], and context vowels [a] and [i]. These were chosen to examine a range of laryngeal settings, from firmly adducted [?], to regular adduction for voiced consonants and vowels, to abducted for voiceless consonants, and are similar to tasks used in clinical settings of vocal-fold mobility like sniffing (fast inhalation), quiet breathing, phonation, and tight adduction (e.g., the Valsalva maneuver). Each VCV utterance was produced five times maintaining a consistent fundamental frequency to minimize changes in larynx height. Imaging used a standard clinical convex probe (Telemed MC4-2R20S-3) with 4 MHz excitation and standard line density. It was hand held by the speaker in transverse orientation at the level of the laryngeal protuberance, with 6 cm depth-of-field sampled at 100 fps, providing 8 pixels/mm resolution (Fig. 1). Audio was co-recorded with the Telemed frame synchronization signal, used in post-processing to align speech audio with the ultrasound video stream, and to generate phone-based forced alignment of each utterance using the Montreal Forced Aligner (McAuliffe et al., 2017). For quantification image sequences were first corrected for probe movement by computing and applying a rotation/translation matrix through comparison with a baseline reference image. A sector mask encompassing the abducted folds was then used to retain only those pixels within the mask (see Fig. 1b for the position of the mask). The log-sum of pixel intensities within the mask was used as a proxy for the extent of adduction for that frame, with higher values (showing more included reflective tissue) indicating greater adduction.

Results: Figure 2 shows the time course of these intensity signals aligned on the acoustic midpoint of the target C for productions from a male speaker. As expected [?] shows the greatest extent of adduction, but systematic differences in the timing and degree of ab/adduction are seen across all C types. Quantification of the timing of devoicing gestures obtained through this approach can improve our understanding of its relationship to acoustic voice onset time (VOT) (Abramson & Whalen, 2017). Ongoing work using trained U-net convolutional models to consistently track image features associated with paired arytenoids has the potential to identify asymmetries in controlling cartilage structures associated with voice pathologies.

Conclusions: Ultrasound imaging offers a useful alternative to videoendoscopic methods for imaging laryngeal kinematics. Quantification of the image sequences it provides has the potential to increase knowledge of how laryngeal structures are controlled not just for phonation, voice quality and fundamental frequency, but also the less studied phenomena of precisely-timed consonant devoicing and the realization of glottal stops.

References:

- Abramson, A.S. & Whalen, D.H. (2017). Voice Onset Time (VOT) at 50: Theoretical and practical issues in measuring voicing distinctions. *Journal of Phonetics*, 63, 75-86.

- Allen, J. E., Clunie, G. M., Slinger, C., Haines, J., Mossey-Gaston, C., Zaga, C. J., Scott, B., Wallace, S. & Govender, R. (2021). Utility of ultrasound in the assessment of swallowing and laryngeal function: a rapid review and critical appraisal of the literature. *International Journal of Language & Communication Disorders*, 56(1), 174–204.
- Coey, C., Esling, J. & Moisik, S. (2019). iPA Phonetics v2.3 (Computer program).
- Matsunaga, A., Niimi, S., & Hirose, H. (1991). Ultrasonic observation of the vertical movement of the larynx during phonation. *Annual Bulletin Research Institute of Logopedics and Phoniatrics*, 25, 47-54.
- McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). Montreal Forced Aligner: Trainable text-speech alignment using Kaldi. Proc. *Interspeech* (2017), 498-502.
- Moisik, S. R., Lin, H., & Esling, J. H. (2014). A study of laryngeal gestures in Mandarin citation tones using simultaneous laryngoscopy and laryngeal ultrasound (SLLUS). *Journal of the International Phonetic Association*, 44, 21-58.
- Poh, D., & Moisik, S. R. (2019). A laryngeal ultrasound study of Singaporean Mandarin tones. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), Proc. 19th ICPHS (Melbourne), 172-176.
- Poletto, C., Verdun, L., Strominger, R. & Ludlow, C. (2004). Correspondence between laryngeal vocal fold movement and muscle activity during speech and nonspeech gestures. *Journal of Applied Physiology*, 97(3), 858-866.
- Pressman, J. & Kelemen, G. (1955). *Physiology of the Larynx*. *Physiological Reviews*, 35(3), 506-554.

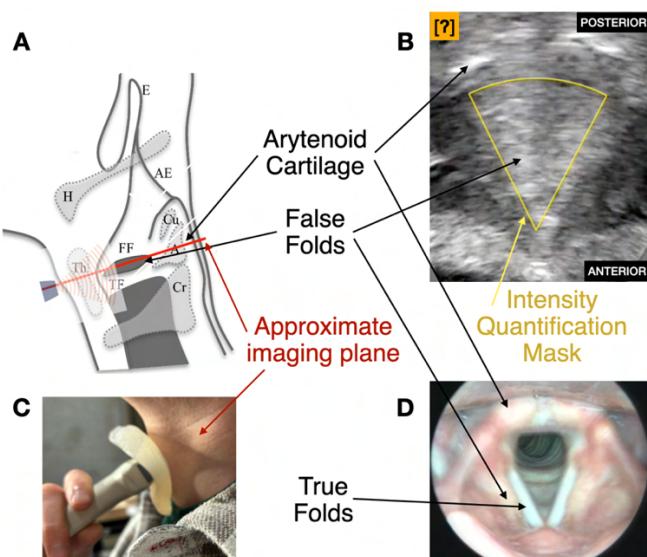


Figure 1. **A)** Orientation and location of transverse ultrasound imaging. **B)** Glottal closure during [?] with superimposed intensity sampling mask in yellow. **C)** Hand-held probe orientation (acoustic standoff in use). **D)** Corresponding structures viewed with endoscopy. (*A and D adapted from Coey et al. (2019)*)

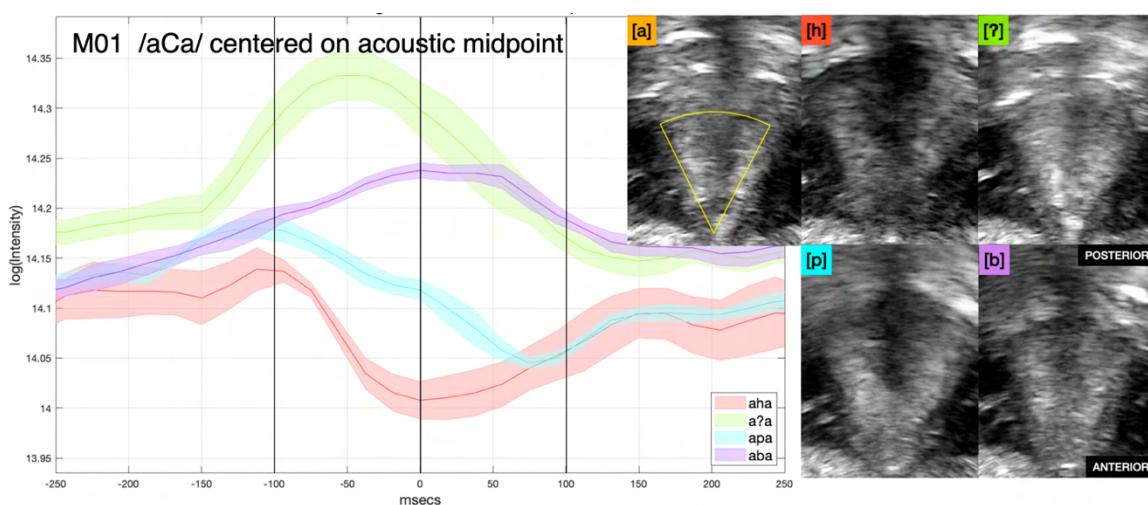


Figure 2. Right side insets show transverse ultrasound example productions by a male speaker at the level of the laryngeal protuberance; each C shown mid-acoustic target. The leftmost [a] image shows the intensity sampling mask in yellow. The false folds are visible at the location of the sides of the mask. Similar positions of the false folds can be seen in the other examples. Note the varying location of the posterior cartilage structures (bright white intensity). The left side plot shows the log sum of pixel values within the sampling mask, given as log(intensity), across frames aligned on the acoustic midpoint of the targeted C. Higher values indicate a greater amount of reflective tissue within the mask associated with greater adduction. Confidence intervals show standard error of the mean (N=5 reps).

Asynchronous Parallel Reinforcement Learning for Voice Control in a Human Larynx Model

Hem Pandeya¹, Xudong Zheng¹, Qian Xue¹

¹ Department of Mechanical Engineering, Rochester Institute of Technology, Rochester, NY, USA

Keywords: Asynchronous Parallel Reinforcement Learning, Voice Control

Abstract

Introduction & rationale: Humans can produce a wide variety of sounds from complex interactions of five intrinsic laryngeal muscles (ILMs), which change the posture of the vocal folds, including their length, thickness, shape, position, and stiffness. The new posture of the vocal folds then interacts with the airflow from the lungs, producing continuous oscillations of vocal folds. These oscillations convert airflow from the lungs into pulsatile airflow, which acts as the main source of sound (Titze, 2000; Zhang, 2016). Each ILM has different functions: the cricothyroid (CT) muscle elongates the vocal folds to increase the pitch, the thyroarytenoid (TA) muscle shortens the vocal folds and adjusts their tension and thickness, the lateral cricoarytenoid (LCA) and interarytenoid (IA) muscles abduct the vocal folds for phonation, and the posterior cricoarytenoid (PCA) muscle abducts them to open the airflow (Luegmair et al., 2014; Yin & Zhang, 2013; Zhang, 2016). Together, these muscles form a highly non-linear system where small changes in the muscle activation level could result in large changes in the acoustic parameters, such as pitch, loudness, and sound pressure level. Because of non-linear behavior, it is difficult to control the acoustic parameters by altering the muscle activation levels in a human larynx model. Precise control of acoustic parameters is important because it can help us to better understand voice mechanics and has potential implications for assisting people with voice disorders and improving voice synthesis.

Several computational approaches have been explored to control the acoustic parameters, such as feedforward-feedback control systems and optimization algorithms (Palaparthi et al., 2024). Among these methods, deep reinforcement learning (DRL) is one of the promising methods because it has the ability to handle non-linear dynamics with high-dimensional state-action spaces (Mnih et al., 2015). The DRL method interacts with the environment and learns an optimal control strategy to achieve a target acoustic profile. However, its trial-and-error nature limits its applications to problems that require high computation power. This issue can be fixed by implementing an asynchronous parallel training (APT) framework, which increases the training efficiency and allows parallel training across heterogeneous hardware.

Objectives:

The objective of this study is to develop an APT-based deep reinforcement learning (DRL) framework integrated with a neuromuscular and fluid-structure-acoustics interaction (FSI) model of the larynx. This framework will enable the prediction of muscle activation strategies for achieving specified flow and acoustic profiles during voice production. Additionally, the study aims to investigate compensatory muscle activation strategies in cases of muscle deficiency.

Methods:

A 3D high-fidelity continuum model of the human larynx was reconstructed from high-resolution MRI scans of a 74-year-old female. The model comprises all of the intrinsic muscles along with all the cartilages (cricoid, arytenoid, and the thyroid). The model incorporates all the major features of the human larynx, such as the cover layer, ligaments, thyroid epiglottis muscle, the conus elasticus, cricoid space, and paraglottic space of adipose tissue (Fig.1(a)). This laryngeal model was fully coupled with a FSI model to simulate vocal fold vibrations for different vocal fold postures. The finite element method was used to obtain the pre-phonatory status of the vocal folds and the one-dimensional Bernoulli equation was solved in order to calculate the glottal flow. From this glottal flow, various acoustic parameters, such as fundamental frequency, sound pressure level, and sound-to-noise ratio, were computed.

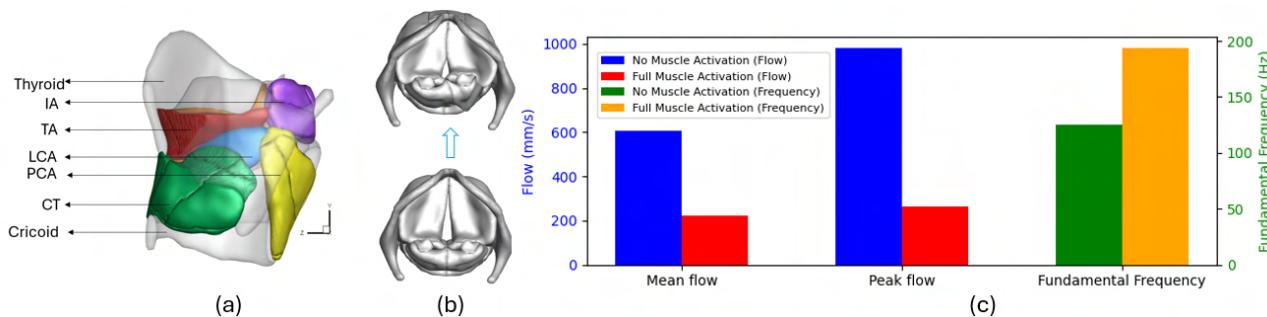


Figure 1: (a) Larynx model showing all five intrinsic muscles. (b) Vocal fold shape in the cadaveric position and after full activation of TA, LCA, and IA muscles, reducing the glottal gap, through muscle activation simulations. (c) Bar chart showing three different acoustic parameters (mean flow, peak flow, and fundamental frequency) at the cadaveric position and full muscle activation through FSI simulations

A DRL agent was assigned to learn an optimal control strategy from its experience of interacting with a simulated FSI environment. An off-policy algorithm was used for the training purpose because of which the past experiences of interaction with the environment were stored in a replay buffer. The APT strategy was implemented to decouple the environment interaction from the neural network training, enhancing computational efficiency. Fig 2 shows the APT strategy, in which CPUs handle all environment interactions without waiting for the GPUs to finish the neural network training epochs. The main advantage of this approach is that it eliminates the waiting time due to the dependence of the CPU on the GPU and vice-versa (Liu et al., 2024).

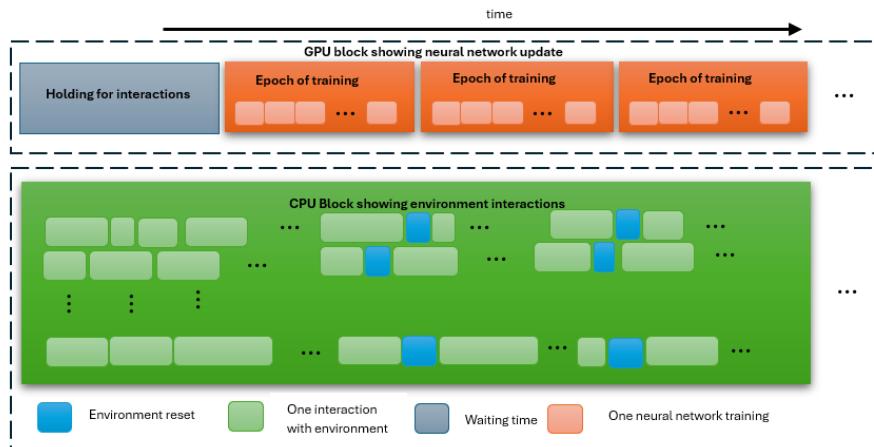


Figure 2: Showing implementation of APT DRL training strategy for reducing computational time

Results: We were able to generate different flow and acoustic quantities by activating the five intrinsic muscles at different levels. For instance, setting the activation level of TA, LCA, and IA muscles to 1 (full activation condition) reduced the glottal gap (Fig 1 (b)). This constriction led to a decrease in both the mean and peak airflow rate while simultaneously increasing the fundamental frequency of the sound (Fig 1(c)).

Next, we will implement this fluid-structure-acoustic interaction model in the DRL framework, where a trained DRL agent will learn control strategies to control the activations of the ILMs to achieve specified flow and acoustic profiles. We will also explore the compensatory muscle activation strategies in cases of muscle deficiency, such as the compensations among TA, LCA and IA in controlling glottis gap.

Conclusions: A high-fidelity three-dimensional model of the human larynx was reconstructed using MRI scan data, which was fully coupled with the FSI model to simulate vocal fold vibrations. An APT framework was implemented on a deep reinforcement learning method to allow parallel training and enhance computational efficiency. This method will be used to train the DRL agent to learn optimal control strategies to achieve the target flow and acoustic profiles, such as frequency against time graph.

References:

- Liu, X.-Y., Bodaghi, D., Xue, Q., Zheng, X., & Wang, J.-X. (2024). Asynchronous parallel reinforcement learning for optimizing propulsive performance in fin ray control. *Engineering with Computers*. <https://doi.org/10.1007/s00366-024-02093-w>
- Luegmair, G., Chhetri, D. K., & Zhang, Z. (2014). The role of thyroarytenoid muscles in regulating glottal closure in an in vivo canine larynx model. 060007. <https://doi.org/10.1121/2.0001504>
- Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., & Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*, 518(7540), 529–533. <https://doi.org/10.1038/nature14236>
- Palaparthi, A., Alluri, R. K., & Titze, I. R. (2024). Deep Learning for Neuromuscular Control of Vocal Source for Voice Production. *Applied Sciences*, 14(2), 769. <https://doi.org/10.3390/app1402076>
- Titze, I. R. (2000). *Principles of Voice Production* (Second). National Center for Voice and Speech.
- Yin, J., & Zhang, Z. (2013). The influence of thyroarytenoid and cricothyroid muscle activation on vocal fold stiffness and eigenfrequencies. *The Journal of the Acoustical Society of America*, 133(5), 2972–2983. <https://doi.org/10.1121/1.4799809>
- Zhang, Z. (2016). Mechanics of human voice production and control. *The Journal of the Acoustical Society of America*, 140(4), 2614–2635. <https://doi.org/10.1121/1.4964509>

Quantifying the Vocal Fold Oscillations using dynamic 3D Magnetic Resonance Imaging

Paula Jordan^{1,3}, Johannes Fischer^{1,3}, Fiona Stritt^{2,3}, Marie Köberlein⁴, Louisa Traser^{2,3}, Bernhard Richter^{2,3}, Michael Bock^{1,3}

¹Dept. of Radiology, Medical Physics, Medical Center University of Freiburg, Freiburg, Germany

²Freiburg Institute of Musicians' Medicine, Medical Center, University of Freiburg, Freiburg, Germany

³Faculty of Medicine, University of Freiburg, Freiburg, Germany

⁴Department of Otorhinolaryngology, Ludwig-Maximilians-Universität München, Division of Phoniatrics and Pediatric Audiology, LMU Klinikum, Munich, Germany

Keywords: Vocal Folds, Magnetic Resonance Imaging, high-speed imaging

Introduction & rationale: A huge variety of measurement techniques has been developed to quantify vocal fold (VF) oscillations parameters such as displacement of the caudal surface and tissue velocity [1–4]. With state-of-the-art optical methods, visualization of the VF anatomy is limited, and quantifying spatial features is challenging. However, a more detailed visualization of the VF 3D motion is necessary to improve the understanding of voice production and VF lesion formation. Recently we showed for the first time that VF oscillation can be imaged using magnetic resonance imaging (MRI) with a high temporal resolution [5,6].

Objectives: In this work, we further developed a dynamic 3D-imaging sequence with sub-millimeter sub-millisecond resolution. This enables the extraction of 3D anatomical data and quantification of the vocal fold oscillations. We aim to extract absolute measurements of vocal fold contact area, as well as quantification of vertical vocal fold motion.

Methods: A zero echo time (ZTE) sequence with a maximum readout gradient strength of $G_1 = 22.8 \text{ mT/m}$ and a data acquisition time of 702 μs was developed for a clinical 3T MRI system (PrismaFit, Siemens). Radial data were acquired using a spiral phyllotaxis pattern [7]. To compensate for an involuntary larynx shift during the measurement additional MR projections were acquired every 375 ms (superior-inferior direction) and 750 ms (left-right and anterior-posterior direction). This allows for retrospective correction of variations in the larynx position relative to a reference position using a phase only cross correlation algorithm [8,9]. To reduce artifacts due to the finite pulse duration ($t = 12 \mu\text{s}$) small residual displacements of the larynx from the isocenter were corrected by adjusting the RF pulse frequency for each projection [10]. To acquire the central volume of the k-space, measurements were interleaved with acquisitions using lower readout gradients ($G_2 = G_1/2$, $G_3 = G_1/4$, $G_4 = G_1/8$, $G_5 = G_1/16$). Data acquisition was repeated every 1.25 ms.

For reception of the MR signal a small circular radio-frequency coil ($d = 7 \text{ cm}$) was placed on the volunteer's larynx. Dynamic 3D MRI data were acquired in 5½ min with 0.77mm spatial resolution. During imaging, the volunteer (male, 35 y, 75 kg) was lying in a supine position and was singing at a constant fundamental frequency ($f_0 = 110 \text{ Hz}$) with intermittent breathing at will. For synchronization between MRI and VF motion, an optical microphone was placed at the mouth of the volunteer recording the acoustic signal. For suppression of the noise from the MRI system, acoustic data were band-pass filtered ($f_c = 80\text{-}230 \text{ Hz}$). Generally, the ZTE sequence is quiet (60 dB) in comparison to the ambient noise (50 dB), and conventional MRI sequences with similar spatial resolution (84 dB). The phase of the VF motion was then obtained from a fit of a sine function for each acquisition. Based on the motion-phase, dynamic data were reconstructed into 10 frames using total variation (TV) constraints along temporal and spatial dimension with the BART software package [11]. TV regularization parameters were determined based on reference values using an adapted version of the sequential S-curve method [12]. A sensitivity map was generated from the data acquired with G_5 to counteract signal-loss with distance from the coil.

Results: The mean phonation frequency was $106 \pm 2 \text{ Hz}$, resulting in a temporal resolution of $943 \mu\text{s}$ per frame (Fig. 1). The SI-displacement of the larynx relative to a reference position was between -2 and 6 mm ($\sigma = 0.2 \text{ mm}$). The coronal view shows a maximum vertical displacement of 1.5 mm of the VF, and transverse images depict a maximum opening width of 4.6 mm (Fig. 2). The VF contact area, extracted by thresholding the dynamic data, shows values of $A_c = 28\text{-}47 \text{ mm}^2$. Furthermore, the glottal area waveform (GAW) was modeled from slices in transverse orientation showing values between 15 and 81 mm^2 (Fig. 3). The MR data show a mean open quotient (OQ) of 0.8 ($\sigma = 0.1$) while EGG data show OQ = 0.6 ($\sigma = 0.04$) while highspeed imaging data of the same volunteer shows OQ = 0.7 ($\sigma = 0.1$).

Conclusions: In this work, VF oscillations were quantified using displacement, opening width, OQ and VF contact area extracted from dynamic MRI data. Results were compared to state-of-the-art methods (EGG and high-speed laryngoscopy). The short encoding time of ZTE is advantageous as motion blurring during acquisition is minimized. The variable pattern of sampling increased the tissue-air-contrast enabling a more detailed quantification of the VF motion. A limitation of the MRI method are the long acquisition times, which might make it unsuitable for patients with VF pathologies.

In voice research, optical imaging of the VF requires the insertion of a laryngoscope into the pharynx. This is challenging for singers and can only provide limited views of the VF anatomy. This limitation can be overcome by dynamic 3D VF MRI, which is providing exact 3D anatomical information to determine quantitative parameters of different periodic phonation types.

Acknowledgments: Grant support from the Deutsche Forschungsgemeinschaft (DFG) under grant numbers FI 2803/1-1 and TR 1491/4-1 is gratefully acknowledged.

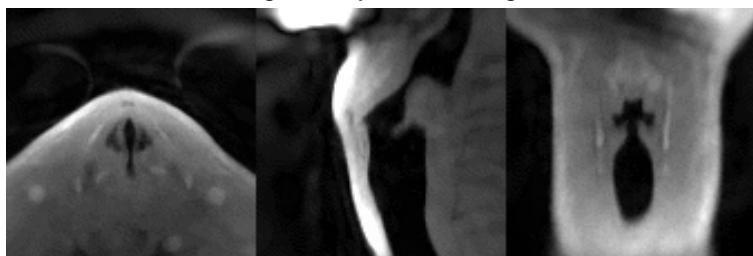


Figure 1: Transverse, sagittal and coronal views of the vocal folds with a spatial resolution of 0.77mm taken from the 3D+t MR data.

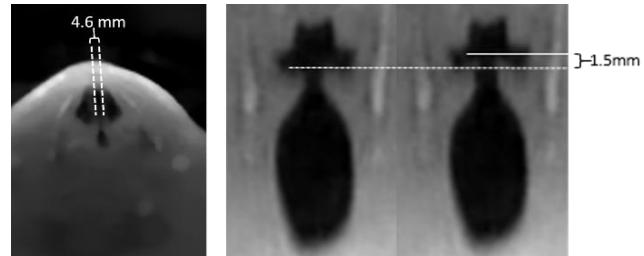


Figure 1: Visualization of the maximum opening width (left) and vertical displacement (right).

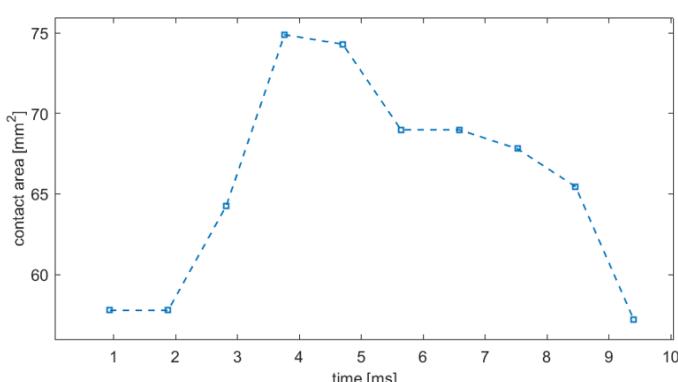
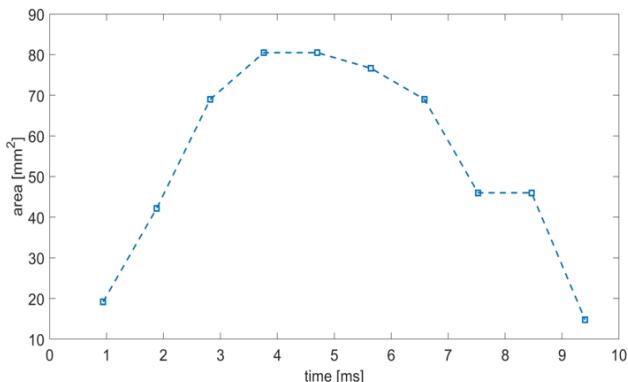


Figure 3: Vocal fold contact area (left) and glottal area waveform (right) extracted from the dynamic MR data using a threshold approach.



References

- Coughlan, C. A., Chou, L., Jing, J. C., Chen, J. J., Rangarajan, S., Chang, T. H., Sharma, G. K., Cho, K., Lee, D., Goddard, J. A., Chen, Z., & Wong, B. J. F. (2016). In vivo cross-sectional imaging of the phonating larynx using long-range Doppler optical coherence tomography. *Scientific Reports*, 6, 22792.
- Švec, J. G., & Schutte, H. K. (1996). Videokymography: High-speed line scanning of vocal fold vibration. *Journal of Voice*, 10(2), 201–205.
- Luegmair, G., Mehta, D. D., Kobler, J. B., & Döllinger, M. (2015). Three-dimensional optical reconstruction of vocal fold kinematics using high-speed video with a laser projection system. *IEEE Transactions on Medical Imaging*, 34(12), 2572.
- Shau, Y.-W., Wang, C.-L., Hsieh, F.-J., & Hsiao, T.-Y. (2001). Noninvasive assessment of vocal fold mucosal wave velocity using color doppler imaging. *Ultrasound in Medicine & Biology*, 27(11), 1451–1460.
- Fischer, J., Abels, T., Özen, A. C., Echternach, M., Richter, B., & Bock, M. (2020). Magnetic resonance imaging of the vocal fold oscillations with sub-millisecond temporal resolution. *Magnetic Resonance in Medicine*, 83(2), 403–411.
- Fischer, J., Özen, A. C., İlbey, S., Traser, L., Echternach, M., Richter, B., & Bock, M. (2022). Sub-millisecond 2D MRI of the vocal fold oscillation using single-point imaging with rapid encoding. *Magnetic Resonance Materials in Physics, Biology and Medicine*, 35(2), 301–310.
- Piccini, D., Littmann, A., Nielles-Vallespin, S., & Zenge, M. O. (2011). Spiral phyllotaxis: The natural way to construct a 3D radial trajectory in MRI. *Magnetic Resonance in Medicine*, 66(4), 1049–1056.
- Reichert, A., Bock, M., Reiss, S., Overduin, C. G., Fütterer, J. J., & Krafft, A. J. (2018). Simultaneous slice excitation for accelerated passive marker tracking via phase-only cross correlation (POCC) in MR-guided needle interventions. *Magnetic Resonance Materials in Physics, Biology and Medicine*, 31(6), 781–788.
- Kuglin, C. D., & Hines, D. C. (1975). *The Phase Correlation Image Alignment Method*. 163–165.
- Ilbey, S., Jung, M., Emir, U., Bock, M., & Özen, A. C. (2022). Characterizing Off-center MRI with ZTE. *Zeitschrift Für Medizinische Physik*.
- Uecker, M., Ong, F., Tamir, J. I., Bahri, D., Virtue, P., Cheng, J. Y., Zhang, T., & Lustig, M. (2015). *BART: Version 0.2.09* [Computer software]. Zenodo.
- Hanhela, M., Gröhn, O., Kettunen, M., Niinimäki, K., Vauhkonen, M., & Kolehmainen, V. (2021). Data-Driven Regularization Parameter Selection in Dynamic MRI. *Journal of Imaging*, 7(2), 38. <https://doi.org/10.3390/jimaging7020038>

POSTER SESSION 2

Quantifying Anatomic, Morphologic, Acoustic, and Perceptual Characteristics of Tracheoesophageal Voice Pre and Post Balloon Dilation Utilizing High Speed Digital Imaging

Angela Dionisio¹, Layne Singer¹, Charles Farbos De Luzan¹, Gregory R. Dion¹

¹Otolaryngology Head and Neck Surgery, University of Cincinnati, Cincinnati, Ohio, USA

Keywords: High speech digital imaging, total laryngectomy, tracheoesophageal voicing, balloon dilation

Introduction & rationale: Although videostroboscopy serves as the gold standard for clinical examination of true vocal fold vibration and analysis of vocal fold pathologies, high speed digital imaging (HSDI) is superior in evaluating vocal fold vibration and identifying functional deficits.(Olthoff et al., 2007) Evaluation tools such as HSDI are important to identify optimal vs suboptimal voicing patterns and thus assist in determining appropriate treatments to optimize the voice. With the expanding role of HSDI in vocal fold pathologies, a critical knowledge gap exists of these HSDI parameters in patients with alaryngeal speech, particularly before and after interventions to improve their phonation. Although foundational work has been performed,(van As et al., 1999) no data exists on HSDI parameters before and after interventions to improve phonation.

Alaryngeal speech occurs in the context of individuals who have undergone total laryngectomy (TL), removal of the larynx. Multiple voice rehabilitation options exist, with tracheoesophageal (TE) speech being the gold standard. TE speech involves the surgical creation of a tracheoesophageal puncture (TEP) and subsequent fitting of a valved silicone prosthesis. Creation of a TEP allows for the re-routing of pulmonary air into the esophagus, vibrating the pharyngoesophageal segment, creating a new source of sound. HSDI can be a useful tool in examining neoglottic vibrations during TEP speech to identify anatomic and morphologic features and measure integrity and characteristics of the neoglottis.(van As-Brooks et al., 2005) Some of these features included assessability, brightness, and focus of the overall recording, amount of saliva, visibility of origin of the neoglottis, shape of the neoglottis, location of visible vibration, level of presence of mucosal wave, regularity of vibration, and closure phase (open, equal, or closed).(van As-Brooks et al., 2005) Additionally, acoustic data can be reliability obtained from TE voices and be a method for obtaining data to objectively measure TE voice quality, with acoustic and perceptual measures that have a moderate to strong correlation.(van As-Brooks et al., 2006)

Patients frequently struggle with a TEP phonation second to a tight neopharynx or fibrosis of the region secondary to radiation therapy and require interventions, frequently with botulinum toxin injections.(Zormeier et al., 1999) Another commonly employed tactic to help with TEP phonation is balloon dilation of the neopharyngeal-esophageal segment, and although patients subjectively report improvements in their voice and effort required to speak, objective data and data on HSDI morphologic and acoustic changes are lacking in this population.

Objectives: This study aims to quantify HSDI of the neoglottis pre and post neopharyngeal - esophageal dilations for TEP phonation to identify anatomical and morphologic factors that may explain the improved phonation after dilation. Coupled with perceptual analyses and subjective data, this information will better elucidate the phonatory mechanics involved in improved TEP phonation after dilation.

Methods: After IRB approval and informed consent, acoustic and HSDI were collected before, and a few weeks after, dilation of the neopharyngeal – esophageal segment. A standardized connected speech sample, “Rainbow Passage” was collected along with sustained /a/ and /i/ sounds (three times for three seconds each). Acoustic pressure was recorded using a Cardioid Condenser Microphone (Audio Technica AT2021) 5 cm from the subject’s mouth, connected to a Tascam US-2x2HR 2 Mic 2IN/2OUT high resolution versatile USB audio interface. Data was acquired via Praat at a sampling rate of 44.1kHz. Acoustic data obtained were reflected in Fundamental frequency (Hz), Jitter (%), Harmonic to noise ratio, Glottal to noise excitation (GNE) ratio, Maximum phonation time (MPT) (s), Cepstral Peak Prominence, mean-Jitt (ms), Jitt-Factor, and Sound pressure level (dB). These parameters were collected through the Glottal Analysis Tool (GAT). HSDI was captured using a transnasal flexible endoscope connected to a high-speed camera system (Photron Nova S9) to capture detailed images. Images were recorded at a frame rate of 4,000 frames per second (fps) with a resolution of 256 × 256 pixels, ensuring precise temporal and spatial resolution of VP dynamics. The high-speed camera was controlled and monitored using Photron’s proprietary software, Fastcam Viewer (PFV). Additionally, a Larson Davis 831 sound pressure level (SPL) meter was employed to measure acoustic baseline parameters prior to data acquisition. Images were imported into Matlab and lines were traced over the major and minor axes of the opening (Fig. 1). A videokymograph (VKG) (Svec, 1999) was then extracted from these axes. The opening area

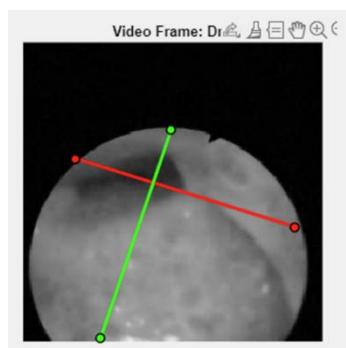


Figure 1. Screenshot of the app used to extract video-kymographs from endoscopic highspeed imaging. The lines are manually traced over the opening main axes.

waveform was derived from the VKG and compared pre and post dilation. Perceptual ratings of the pre and post balloon dilation connected speech samples were then completed using the Sunderland Tracheoesophageal Perceptual Scale (SToPs). (D'Alatri et al.) Voice samples were rated by two expert raters, who were blinded to the condition (pre vs post dilation). Perceptual parameters include overall severity rating, perceptual tonicity, strain (amount of audible effort), level of wetness/gurgliness, impairment of volume, social acceptability, perceptual perception of amount of whisper, intelligibility, stoma noise/burst (level of noise competing with oral speech), and speech fluency.

Results: Video-kymographs were compared before and after dilation for both vowels (**Fig. 2**) with post-dilation kymographs revealing higher intensity peaks and increased mucus fluctuations compared to pre-dilation. This observation aligns with our hypothesis that dilating the TEP valve would allow more mucus secretions by increasing the available volume. Average peak heights of the derivative of intensity (representative of vocal fold closing speed) for pre-dilation and post-dilation conditions demonstrate post-dilation conditions with higher average peak heights, suggesting faster closing speeds (**Fig. 3**). Data across all vowels highlights an overall increase in closing speed post-dilation, consistent with the hypothesis that TEP valve dilation enhances mucus dynamics and mucosal wave activity. The data suggest that the closing speed is higher for post /a/ compared to pre /a/, while the trend is reversed for /i/. Overall, the much higher peaks for post /a/

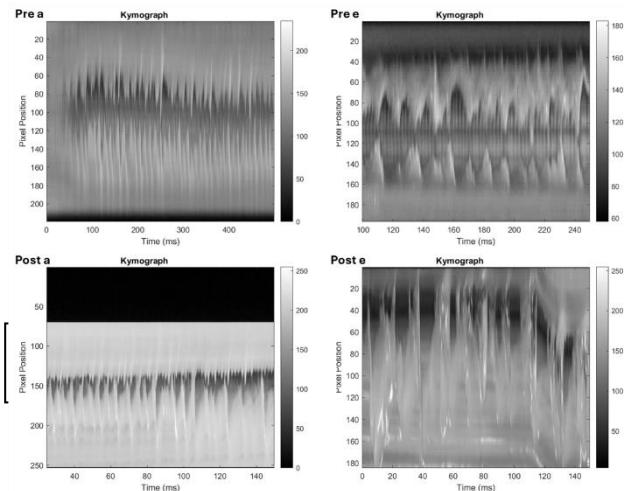


Figure 2. Video-kymographs for pre-dilation (top), compared to post dilation (bottom), and for /a/ (left) and /i/ (right).

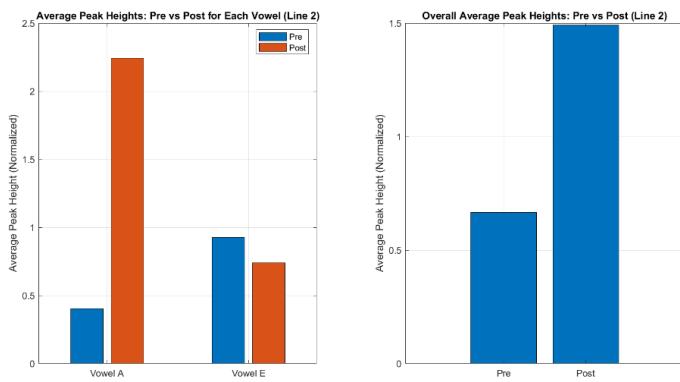


Figure 3. Closing speed analogy based on video-kymograph analysis.

Conclusions: Acoustic and neoglottic morphologic changes were quantified before and after balloon dilation in a patient with subjective voice improvement. These observations illustrate changes that occur with dilation to improve voice, which target characteristics for evaluation in an expanded patient population. Additional insights on subjective and perceptual voice quality ratings are needed to improve TEP voice rating mechanisms.

References:

- D'Alatri, L., Longobardi, Y., Marchese, M. R., Meschini, C., Figliolia, D., Mari, G., Rodolico, D., Parrilla, C., & Galli, J. Perceptual Evaluation of the Tracheoesophageal Voice: Italian Validation of the Sunderland Tracheoesophageal Perceptual Scale. *Journal of Voice*. <https://doi.org/10.1016/j.jvoice.2023.11.007>
- Olthoff, A., Woywod, C., & Kruse, E. (2007). Stroboscopy Versus High-Speed Glottography: A Comparative Study. *The Laryngoscope*, 117(6), 1123-1126. <https://doi.org/10.1097/MLG.0b013e318041f70c>
- van As-Brooks, C. J., Hilgers, F. J. M., Koopmans-van Beinum, F. J., & Pols, L. C. W. (2005). Anatomical and Functional Correlates of Voice Quality in Tracheoesophageal Speech. *Journal of Voice*, 19(3), 360-372. <https://doi.org/10.1016/j.jvoice.2004.07.011>
- van As-Brooks, C. J., Koopmans-van Beinum, F. J., Pols, L. C. W., & Hilgers, F. J. M. (2006). Acoustic Signal Typing for Evaluation of Voice Quality in Tracheoesophageal Speech. *Journal of Voice*, 20(3), 355-368. <https://doi.org/10.1016/j.jvoice.2005.04.008>
- van As, C. J., Tigges, M., Wittenberg, T., Op de Coul, B. M. R., Eysholdt, U., & Hilgers, F. J. M. (1999). High-Speed Digital Imaging of Neoglottic Vibration After Total Laryngectomy. *Archives of Otolaryngology–Head & Neck Surgery*, 125(8), 891-897. <https://doi.org/10.1001/archotol.125.8.891>
- Zormeier, M. M., Meleca, R. J., Simpson, M. L., Dworkin, J. P., Klein, R., Gross, M., & Mathog, R. H. (1999). Botulinum toxin injection to improve tracheoesophageal speech after total laryngectomy. *Otolaryngol Head Neck Surg*, 120(3), 314-319. [https://doi.org/10.1016/s0194-5998\(99\)70268-8](https://doi.org/10.1016/s0194-5998(99)70268-8)

result in a higher average closing speed post-dilation. Post-dilation, there were also notable changes in MPT, fundamental frequency, glottal-to-excitation ratio, and Jitt-factor. The MPT decreased from 7.2 seconds to 5.3 seconds. This observation shows there is less obstruction for expired air, leading to a decrease time the subject can sustain phonation. Furthermore, there was an overall decrease seen in both GNE ratio and Jitt-factor. Lastly, there was an overall increase in fundamental frequency with /i/ that was not seen in the /a/ phonation. Interestingly, perceptual voice ratings decreased after dilation.

Measuring the effects of corticosteroid use on vocal fold inflammation and biomechanics using a rabbit model

Maya Stevens^{1,2}, Brendan Olson³, Anika Isom⁴, Douglas Roberts⁵, Jenny Pierce², Julie M Barkmeier-Kraemer^{1,2}, Ray Merrill⁶, Kristine Tanner¹, Ben Christensen^{2,3}

¹Department of Communication Sciences & Disorders, University of Utah, SLC, UT, USA

²Department of Otolaryngology – Head & Neck Surgery, University of Utah School of Medicine, SLC, UT, USA

³Department of Surgery, Division of Urology, University of Utah School of Medicine, SLC, UT, USA

⁴Department of Biology, University of Utah, SLC, UT, USA

⁵Department of Bioengineering, University of Utah, SLC, UT, USA

⁶Department of Communication Sciences & Disorders, Brigham Young University, Provo, UT, USA

Keywords: Asthma; Dysphonia; Inflammation; Elastic Modulus

Abstract

Introduction & rationale: Approximately 8% of the population has a diagnosis of asthma, with the prevalence increasing each year. A commonly prescribed treatment for asthma is combination inhaled corticosteroids (ICs), which are composed of corticosteroids and beta agonists that relax the smooth muscle of the airway. Unfortunately, the use of combination ICs has been associated with voice disorders in 5-58% of patients¹. This association has been attributed to the fact that the drugs must pass through the airway before reaching the lungs.

Objectives: The purpose of this study is to investigate the mechanisms that may be associated with voice disorders after IC usage by determining the inflammatory and biomechanical changes in vocal fold tissue following IC administration in a rabbit model.

Methods: For Phase 1 of this study, 20 6-month old male New Zealand white rabbits were randomly assigned to a control or experimental group (N=10 each). The experimental group received Advair HFA, a combination corticosteroid consisting of fluticasone propionate (45 mcg) and salmeterol (21 mcg). Medication was administered via an inhaler with spacer and mask. The control group received a saline treatment with 0.9% sodium chloride via a nebulizer with a small animal mask attachment. Both groups were administered 18 breaths of one puff or nebulization twice daily for eight consecutive weeks. Following the treatment period, all animals were euthanized, and vocal folds were harvested and stored in phosphate-buffered saline (PBS) at -80°C. The right vocal folds were thawed prior to rheometric tests to determine elastic and viscous moduli. Following rheometry, RNA was extracted from the tissues and processed for RT-qPCR analysis. The tissues were tested for the following pro-inflammatory cytokines: TNF-α, IL-1β, and IL-6.

For Phase 2 of this study, rabbits were divided into five groups: baseline control; induction (rabbits were euthanized when visual perceptual changes to the vocal folds occurred based on expert rater evaluation of edema and erythema); induction control (paired controls for the induction group which received saline); reversibility (rabbits were administered medication until visual changes in the vocal folds, and then withdrawn from medication and euthanized when visual changes returned to baseline); and reversibility control (paired controls for the reversibility group). Medication and saline were administered as in Phase 1, and rheology and PCR analysis were also similarly conducted.

Results: For Phase 1, viscoelastic differences were not identified between vocal folds treated with corticosteroids and those treated with saline. Although significant differences were not found in TNF-α and IL-1β expression, IL-6 expression was significantly increased in the combination IC treated group.

For Phase 2, moduli for the induction group were higher (stiffer) than induction controls and baseline values. Moduli for reversibility and reversibility controls were also higher than induction controls and baseline values, but similar to each other.

Conclusions: For Phase 1, findings indicated an increased amount of IL-6 in vocal folds treated with corticosteroids compared to control vocal folds. IL-6 is an indicator of chronic inflammation, and suggests that voice disorders due to IC treatment may be a consequence of a chronic inflammatory response to the medication. For Phase 2, moduli for the induction group was higher than control and baseline values, suggesting that the confirmed presence of inflammation in the vocal folds based on visual ratings does lead to biomechanical changes.

References:

1. Galvan CA, Guarderas JC. Practical considerations for dysphonia caused by inhaled corticosteroids. *Mayo Clin Proc*. Sep 2012;87(9):901-4. doi:10.1016/j.mayocp.2012.06.022

Development and validation of personal technology to treat hypophonia in patients with Parkinsons Disease

Amanda I Gillespie¹, Adam Klein¹, Meredith Caveney², David Anderson²

¹ Emory University School of Medicine, Atlanta, GA, USA

² Georgia Institute of Technology, Atlanta, GA, USA

Keywords: technology, Parkinsons, voice

Abstract

Introduction & rationale: Parkinson's disease (PD) severely impacts communication in 70-90% of patients. Behavioral voice therapy can be effective in Parkinson's-related voice and speech disorders by increasing vocal loudness and improving speech intelligibility. However, success in treatment is dependent on an intensive intervention schedule. This treatment burden, combined with reduced ability of patients with Parkinson's to self-monitor voice leads to a high rate of symptomatic relapse following treatment.

Objectives: The purpose of this study was to develop and validate wearable technology to improve vocal intensity in patients with PD

Methods: Our multi-disciplinary, multi-institutional team created a low-profile headset device that can isolate, monitor, and analyze vocal output for time and loudness and provide haptic biofeedback in the form of vibration to the wearer when minimum threshold intensity targets are not met. In this study, we tested the device in 8 patients with PD during a 60-minute treatment session. Each session consisted of 30 minutes with and 30 minutes without speech pathologist's feedback on vocal intensity. Intensity change and time between cues to increase intensity were measured and analyzed for differences between baseline and each condition.

Results: There were no significant differences ($p = 0.25$) in time between cues to increase intensity in the SLP assisted condition compared to the device condition. There was no significant difference for change in vocal intensity in decibels between the SLP assisted condition and the device assisted condition ($p = 0.66$). When groups were split by disease severity, those with mild to moderate PD required significantly fewer intensity cues and increased vocal intensity significantly more than those with severe PD for both the SLP and device-assisted conditions ($p < .001$).



Conclusions: Patients with mild to moderate PD respond positively to the haptic feedback provided by an individually calibrated low profile headset. This technology may serve as a helpful adjunct to standard-of-care voice therapy. Future studies will investigate if the addition of the device in voice therapy improves outcomes including a reduction in treatment time and relapse.

References:

- Clark JP, Adams SG, Dykstra AD, Moodie S, Jog M. (2014) Loudness perception and speech intensity control in Parkinson's disease. *Journal of communication disorders*. 51:1-12.
- Ramig LO, Fox C, Sapir S. (2004). Parkinson's disease: speech and voice disorders and their treatment with the Lee Silverman Voice Treatment. *Seminars in speech and language*. 25(2):169-180.
- Gustafsson J, Ternström S, Södersten M, Schalling E. (2016) Motor-Learning-Based Adjustment of Ambulatory Feedback on Vocal Loudness for Patients With Parkinson's Disease. *Journal of voice*. 30(4):407-415.
- Schalling E, Gustafsson J, Ternström S, Bulukin Wilén F, Södersten M. (2013) Effects of tactile biofeedback by a portable voice accumulator on voice sound level in speakers with Parkinson's disease. *Journal of voice*. 27(6):729-737.

MODELING UNCERTAINTY IN SPEECH ADAPTATION TO SENSORY FEEDBACK PERTURBATIONS

Nicolás F. Quinteros^{1,2}, Juan P. Cortés², Jesús A. Parra², Hasini R. Weerathunge³, Frank H. Guenther⁴, Matías Zañartu^{1,2}

¹ Department of Electronic Engineering, Universidad Técnica Federico Santa María, Valparaíso, Chile

² Advanced Center for Electrical and Electronic Engineering, Universidad Técnica Federico Santa María, Valparaíso, Chile,

³ Department of Psychiatry, University of Michigan, Michigan

⁴ Department of Speech, Language, and Hearing Sciences and Department of Biomedical Engineering, Boston University, Boston, Massachusetts.

Keywords: Speech Motor Control, Adaptive Neurophysiological Modeling, Stochastic Variability

Introduction & rationale: Altered auditory feedback experiments, commonly used to study the adaptability of speech and voice production in response to auditory perturbations, provide insights into the direction of the response (compensation or following the perturbation), the magnitude of adaptation (partial or complete), and the speed of adaptation related to feedforward learning rates (Guenther, 2016). Kearney et al. (2020) proposed a simple three-parameter model named SimpleDIVA, to analyze adaptation in speech and voice production with estimates of sensory feedback gains (auditory and somatosensory) and a feedforward learning rate. This model offers physiological interpretability to characterize individual and group-level responses to sensory perturbations. However, although the standard error of the mean (SEM) can be calculated from the data, SimpleDIVA primarily fits the mean response at the group level and does not account for variability in the measured production response. This variability is likely a result of the complex interaction between different speech production mechanisms, including feedback and feedforward control. We hypothesize that estimating uncertainty in both production and perception can provide valuable information about the acuity of these mechanisms at the group level (e.g., to characterize groups of individuals with typical and hyperfunctional voices). To this end, we propose to extend the SimpleDIVA model by introducing one or two additional parameters that explicitly represent variability in different speech mechanisms.

Objective: To characterize group response variability in sensory perturbation experiments by extending a simple neurophysiological model of speech adaptation with additional parameters that explicitly capture uncertainty in production and perception processes.

Methods: The dataset (Haenchen, 2017; Scott et al., 2019) consisted of a 60-trial first formant upshift (F1) adaptive paradigm experiment for 10 subjects. The first 19 trials served as the baseline, trials 21 to 40 constituted the hold phase, and trials 41 to 60 were assigned as the after-effect phase. The learning rate and the auditory and somatosensory gains were obtained for the group responses. Subsequently, a modified version of the SimpleDIVA model was implemented considering the three previously obtained parameters and expanded with one and two additional parameters representing uncertainty in production and perception processes of F1. These variability parameters were modeled as standard deviations (SDs) derived from zero-mean white Gaussian noise distributions:

- Sensory perception uncertainty: $\eta_{FB} \sim \mathcal{N}(0, \sigma_{FB}^2)$
- Production uncertainty: $\eta_{Prod} \sim \mathcal{N}(0, \sigma_{Prod}^2)$

These parameters were integrated into SimpleDIVA, where the uncertainty components were incorporated into the SimpleDIVA original equations respectively. The model equations were adapted as follows:

1. Feedback control update: $\Delta F1_{FB}[k] = \alpha_{aud} * (F1_T - F1_{AF}[k] + \eta_{FB}[k]) + \alpha_{som} * (F1_T - F1_{SF}[k] + \eta_{FB}[k])$
2. Feedforward update: $F1_{FF}[k+1] = F1_{FF}[k] + \lambda_{FF} * \Delta F1_{FB}[k]$
3. F1 produced: $F1_{Prod}[k] = F1_{FF}[k] + \Delta F1_{FB}[k] + \eta_{Prod}[k]$

These modifications allow the model to account for variability observed in speech motor control processes, allowing simulations to approximate experimental data.

To assess uncertainty, three cases were analyzed: (1) uncertainty in perception with perfect production acuity, (2) uncertainty in production with perfect perception acuity, and (3) uncertainty in both perception and production. For each case, a Particle Swarm Optimization (PSO) algorithm was employed to optimize model parameters by minimizing the root-mean-square error (RMSE) between the SEM of the experimental data and the simulated data. Optimization was performed using 1,000 particles over 100,000 Monte Carlo simulations per iteration, with a tolerance criterion to ensure convergence. The resulting SDs for each variability parameter were estimated to explain group-level variability effectively.

Results: The SEM produced calculated from the dataset was 31.06 Hz, and the SimpleDIVA parameters for auditory feedback gain, somatosensory feedback gain, and feedforward learning rate were 0.23, 0.17, and 0.11, respectively. Figure 1 illustrates the mean and SEM of Monte Carlo simulations for feedback, feedforward, and production processes, showing that as the number of simulations tends to infinity, the mean of the simulated data converges to the original SimpleDIVA model fit, while the SEM remains constant across trials, with values of 30.40 Hz for feedback difference, 10.40 Hz for feedforward, and 30.51 Hz for production (regardless of whether noise is introduced in a single mechanism or multiple mechanisms, PSO optimization across a large number of Monte Carlo simulations consistently results in the same SEM values for feedback difference, feedforward, and production). However, because noise propagates through the model equations, the SDs of the noise parameters obtained through PSO, whether for one or two parameters, do not translate directly into the SEM values of the simulations directly observed in Figure 1, which is also evident in Table 1, where different noise conditions yield varying SD values, yet the SEM remains consistent when the number of Monte Carlo simulations is sufficiently large. Table 1 also highlights that when only production noise is assumed, the resulting SEM (31.07 Hz) closely matches the SEM obtained from the experimental data (31.06 Hz), while when only perception SEM is considered, there is greater uncertainty (106.35 Hz) in achieving the same variability in the experimental production response (31.06 Hz), mainly due to the low auditory and somatosensory gain values obtained from the evaluated group.

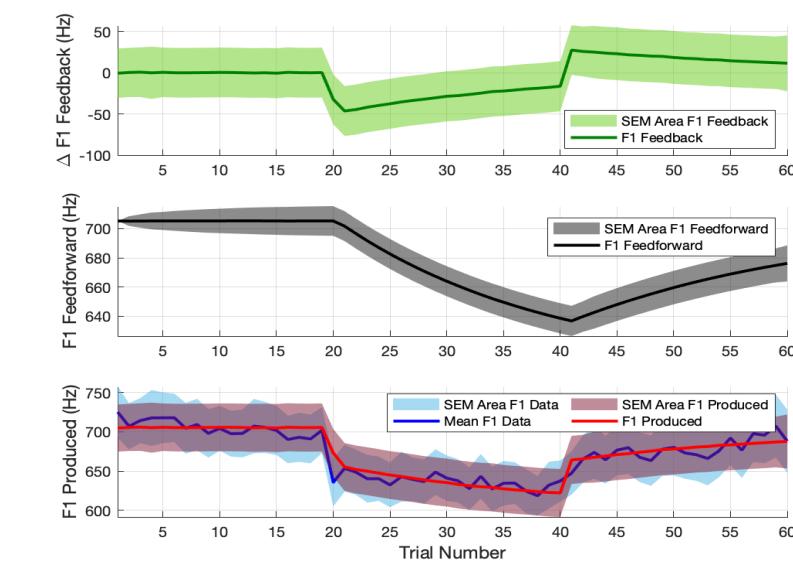


Figure 1: Mean and SEM of 100,000 SimpleDIVA Monte-Carlo simulations with 2 additional noise parameters to characterize the uncertainty of the production data.

Table 1: SD parameters, respective SEM parameters and RMSE metric for different cases of uncertainty.

Case	Perception SD (Hz)	Production SD (Hz)	Perception SEM (Hz)	Production SEM (Hz)	RMSE (Hz)
Perception uncertainty only	330.66	0	106.35	0	2.13
Production uncertainty only	0	96.59	0	31.07	2.43
Uncertainty of perception and production	329.51	5.19	105.98	1.67	2.46

Conclusions: Adding one or two noise parameters into SimpleDIVA to evaluate altered auditory feedback experiments enhances our understanding of how sensory feedback, feedforward, and production mechanisms contribute to group-based uncertainty in experimental data. This not only highlights the variabilities of these mechanisms in achieving SEM of experimental production data given the model, but also demonstrates how the degree of uncertainty in one mechanism can be inferred from the knowledge or assumed acuity of another. For instance, our analysis shows that to achieve the same variability in production data (31.06 Hz, SEM calculated from the experimental data), uncertainty can be attributed entirely to Production (31.07 Hz, SEM as a model parameter), entirely to Perception (106.35 Hz, SEM as a model parameter), or distributed between both (1.67 Hz and 105.98 Hz, respectively, both SEM as model parameters). This approach strengthens our ability to distinguish between different populations (e.g., typical vs. those with speech or voice disorders), offering valuable insights into their unique adaptation mechanisms.

Acknowledgements: This research was supported by the National Institutes of Health (NIH) National Institute on Deafness and Other Communication Disorders grant P50 DC015446, ANID grants FONDECYT 1230828, BASAL AFB240002, and National PhD Scholarship 21240471.

References:

- Guenther, F. H. (2016). *Neural control of speech*. Mit Press.
- Kearney, E., Nieto-Castañón, A., Weerathunge, H. R., Falsini, R., Daliri, A., Abur, D., Ballard, K. J., Chang, S.-E., Chao, S.-C., Heller Murray, E. S., Scott, T. L., & Guenther, F. H. (2020). A simple 3-parameter model for examining adaptation in speech and voice production. *Frontiers in psychology*, 10, 2995.
- Haenchen, L. (2017). *Noninvasive neurostimulation of sensorimotor adaptation in speech production* (Master's thesis, Boston University).
- Scott, T. L., Haenchen, L., Daliri, A., Chartove, J., Guenther, F. H., & Perrachione, T. K. (2019). Speech motor adaptation during perturbed auditory feedback is enhanced by noninvasive brain stimulation. *Trials*, 1, 0-30.

Identifying the Speech Breathing Kinematics of Older Adults with Presbyphonia

Brian Saccente-Kennedy^{1,2}, Jessica Huber³, Martin Wiegand⁴, Matthew Maddocks^{2*}, Roganie Govender^{5*}

¹ Department of Speech and Language Therapy (ENT), Royal National ENT and Eastman Dental Hospitals, University College London Hospitals NHS Foundation Trust, London, UK.

² Cicely Saunders Institute of Palliative Care, Policy and Rehabilitation, King's College London, UK

³ Department of Communication Disorders and Sciences, University at Buffalo, Buffalo, NY, USA.

⁴ Department of Statistical Science, University College London, UK

⁵ Head and Neck Academic Centre, Division of Surgery & Interventional Science, University College London, London, UK.

*Joint last authors

Keywords: presbyphonia, respiratory, kinematics

Abstract

Introduction & rationale:

Although the prevalence of dysphonia in adults over 65 years of age is not yet well established, recent meta-analysis of published data indicates that it is at least twice that of the overall population (Wang et al., 2023). For people in care homes, it is perhaps more than four times as high (Pernambuco et al., 2017). Voice clinic referrals for older adults already outpace their proportion in the population, comprising >20% of dysphonia diagnoses (Davids et al., 2012; Takano et al., 2009). Dysphonia due to age-related changes to the larynx and respiratory system, called *presbyphonia*, is the most common form of dysphonia for adults over 65 (Davids et al., 2012; Jacks et al., 2021).

Research into the pathogenesis and treatment of presbyphonia has largely focused on laryngeal atrophy, *presbylarynx*, with relatively scant attention paid to the role of ageing respiration (Belsky et al., 2021; Desjardins et al., 2020; Saccente-Kennedy et al., 2024). What's more, *presbylarynx* alone has been shown to be insufficient to predict the presence of presbyphonia (Crawley et al., 2018). Thus, the present study seeks to characterise the respiratory presentation of people with presbyphonia, with particular emphasis on speech-related respiratory function, identify how these factors differ from published data for non-dysphonic elderly (Hoit et al., 1989; Hoit & Hixon, 1987; Huber & Spruill, 2008), and determine how they correlate with the subjective experience of vocal handicap. Speech breathing patterns have not yet been described in this population.

Objectives:

- To characterise the speech-breathing patterns of people aged over 50 presenting with presbyphonia.
- To compare these patterns with published normative respiratory kinematic data.
- To identify associations between these patterns and lung function, respiratory muscle strength, vocal handicap, laryngeal aerodynamics and acoustic voice metrics.

Methods:

17 treatment-seeking adults with presbyphonia will be identified from the voice and laryngology clinics of 6 London hospitals, selected for their highly specialised voice services. After passing screening for typical hearing (40dB at 0.5, 1, 1.5 kHz) and absence of cognitive impairment (RUDAS score >22), participants will undergo a battery of voice and respiratory assessments during a single 2-hour research visit. These comprise acoustic assessment (Acoustic Voice Quality Index (AVQI) and Acoustic Breathiness Index (ABI), mean sound pressure level ($SPL_{30\text{ cm}}$)), laryngeal aerodynamics (estimated subglottic pressure, airflow during voicing, laryngeal resistance), speech-breathing assessment at comfortable and loud conditions with inductive respiratory plethysmography (lung volume initiation (LVI), lung volume termination (LVT), ribcage contribution (RCC), %vital capacity per syllable (%VC_{syl}), rate of speech (RoS)), spirometry, respiratory muscle strength (maximum expiratory pressure (MEP), maximum inspiratory pressure (MIP)) and patient-reported outcome measures (Aging Voice Index (AVI), (Etter et al., 2018)). Speech breathing outcomes will be analysed by a single-sample, two-sided t-test for differences in means compared against published data. We will investigate how secondary outcomes impact speech breathing outcomes with a linear regression model.

Results:

This study is ongoing. 5 adults have been screened and recruited to date (31 January 2025), 3 white females (mean age: 81.7) and 2 white males (mean age 75.5). All but one participant had normal spirometry. For respiratory muscle strength, the participants were qualitatively weaker in terms of expiratory muscle strength (median % predicted MEP: 58%) than inspiratory muscle strength (median % predicted MIP: 94%). The median AVI score was 29 [12, 54], which was somewhat lower (more mild) than the mean score found for *presbylarynx* participants in the tool's validation study: 37.8 (Etter et al., 2018).

The median AVQI was 2.87 [1.23, 3.77]; this was above the normative threshold for 80% of participants, indicating disordered voice quality. Median ABI was 3.81 [1.52, 5.42]; this was above the normative threshold for 60% of participants. The median laryngeal resistance was 201 cm H₂O/(l/s) [143, 328]; this was higher than age- and sex-based published norms (Zraick et al., 2012) for 100% of participants.

For monologue, the median intensity in the comfortable loudness condition was 64db [61, 66]. Median LVI (relative to end expiratory level) was 7.2%VC [-2.89, 19.7], median LVT was 1.39%VC [-14.1, 6.5], median RCC was 54.5% [36.5, 73.2], median %VC_{syl} was 0.98%VC [0.51, 1.46] and median RoS was 5.2 syllables/sec [4.7, 6.3]. LVI, RCC and %VC_{syl} were all qualitatively low compared to published data for non-dysphonic older adults (Hoit et al., 1989; Hoit & Hixon, 1987).

For the loud monologue condition, the median intensity was 76dB [73, 80]. Median LVI was 10.1%VC [3.7, 20.1], median LVT was -0.3%VC [-7.8, 5.8], median RCC was 53.7% [30.7, 63.9], median %VC_{syl} was 1.32% [0.68, 1.81] and median RoS was 5.2 syllables/sec [4.5, 5.6]. LVI, LVT were qualitatively low compared to published data for non-dysphonic older adults in a similar loud condition (Huber & Spruill, 2008).

Conclusions:

Data collection is currently ongoing. However, participants are exhibiting low LVI and low LVT compared to published data, which may play a role in the development of presbyphonia and may be targeted with tailored voice therapy interventions. We expect recruitment and preliminary data analysis will be completed in time for presentation at AQL.

References:

- Belsky, M. A., Shelly, S., Rothenberger, S. D., Ziegler, A., Hoffman, B., Hapner, E. R., Gartner-Schmidt, J. L., & Gillespie, A. I. (2021). Phonation Resistance Training Exercises (PhoRTE) With and Without Expiratory Muscle Strength Training (EMST) For Patients With Presbyphonia: A Noninferiority Randomized Clinical Trial. *Journal of Voice*. <https://doi.org/10.1016/j.jvoice.2021.02.015>
- Crawley, B. K., Dehom, S., Thiel, C., Yang, J., Cragoe, A., Mousselli, I., Krishna, P., & Murry, T. (2018). Assessment of clinical and social characteristics that distinguish presbylaryngis from pathologic presbyphonia in elderly individuals. *JAMA Otolaryngology - Head and Neck Surgery*, 144(7), 566–571. <https://doi.org/10.1001/jamaoto.2018.0409>
- Davids, T., Klein, A. M., & Johns, M. M. (2012). Current dysphonia trends in patients over the age of 65: Is vocal atrophy becoming more prevalent? *Laryngoscope*, 122(2), 332–335. <https://doi.org/10.1002/LARY.22397>
- Desjardins, M., Halstead, L., Simpson, A., Flume, P., & Bonilha, H. S. (2020). The Impact of Respiratory Function on Voice in Patients with Presbyphonia. *Journal of Voice*. <https://doi.org/10.1016/j.jvoice.2020.05.027>
- Etter, N. M., Hapner, E. R., Barkmeier-Kraemer, J. M., Gartner-Schmidt, J. L., Dressler, Iie. V., Stemple, J. C., & Carolina, N. (2018). Aging Voice Index (AVI): Reliability and Validity of a Voice Quality of Life Scale for Older Adults. *Journal of Voice*. <https://doi.org/10.1016/j.jvoice.2018.04.006>
- Hoit, J. D., & Hixon, T. J. (1987). Age and Speech Breathing. *Journal of Speech, Language, and Hearing Research*, 30(3), 351–366. <https://doi.org/10.1044/jshr.3003.351>
- Hoit, J. D., Hixon, T. J., Altman, M. E., & Morgan, W. J. (1989). Speech Breathing in Women. *Journal of Speech, Language, and Hearing Research*, 32(2), 353–365. <https://doi.org/10.1044/jshr.3202.353>
- Huber, J. E., & Spruill, J. (2008). Age-Related Changes to Speech Breathing With Increased Vocal Loudness. *Journal of Speech, Language, and Hearing Research*, 51(3), 651–668. [https://doi.org/10.1044/1092-4388\(2008/047\)](https://doi.org/10.1044/1092-4388(2008/047))
- Jacks, A., Kavookjian, H., & Kraft, S. (2021). Comparative Evaluation and Management of Dysphonia Between Adults <65 and ≥65 Years of Age. *Otolaryngology - Head and Neck Surgery (United States)*, 165(1), 142–148. <https://doi.org/10.1177/0194599820978435>
- Pernambuco, L., Espelt, A., Góis, A. C. B., & de Lima, K. C. (2017). Voice Disorders in Older Adults Living in Nursing Homes: Prevalence and Associated Factors. *Journal of Voice*, 31(4), 510.e15–510.e21. <https://doi.org/10.1016/j.jvoice.2016.11.015>
- Saccente-Kennedy, B., Gillies, F., Desjardins, M., Van Stan, J., & Govender, R. (2024). A Systematic Review of Speech-Language Pathology Interventions for Presbyphonia Using the Rehabilitation Treatment Specification System. *Journal of Voice*, S089219972300396X. <https://doi.org/10.1016/j.jvoice.2023.12.010>
- Takano, S., Kimura, M., Nito, T., Imagawa, H., Sakakibara, K.-I., & Tayama, N. (2009). Clinical analysis of presbylarynx-Vocal fold atrophy in elderly individuals. *Auris Nasus Larynx*, 37, 461–464. <https://doi.org/10.1016/j.anl.2009.11.013>
- Wang, L.-H., Doan, T.-N., Chang, F.-C., To, T.-L., Ho, W.-C., & Chou, L.-W. (2023). Prevalence of Voice Disorders in Older Adults: A Systematic Review and Meta-Analysis. *American Journal of Speech-Language Pathology*, 32(4), 1758–1769. https://doi.org/10.1044/2023_AJSLP-22-00393
- Zraick, R. I., Smith-Olinde, L., & Shotts, L. L. (2012). Adult normative data for the KayPENTAX phonatory aerodynamic system model 6600. *Journal of Voice*, 26(2), 164–176. <https://doi.org/10.1016/j.jvoice.2011.01.006>

Differences in Vocal Fold Vibratory Kinematics Between Women with Phonotrauma and Vocally Healthy Controls

Laura E. Toles¹, Ted Mau¹

¹ Department of Otolaryngology – Head & Neck Surgery, University of Texas Southwestern Medical Center, Dallas, TX, USA

Keywords: Phonotrauma; High-speed videoendoscopy

Abstract

Introduction & rationale: The etiology of phonotrauma (e.g., nodules, polyps) is theorized to develop, in part, from increased vocal fold contact pressure and hyperadduction of the vocal folds during phonation (Hillman et al., 2020; Titze et al., 2003). It is a common disorder that can recur or persist over time (Lee et al., 2021), which is why efforts to improve early identification of behaviors leading to phonotraumatic lesions before they develop are important. High-speed videoendoscopy provides a reliable way to quantify vocal fold vibratory patterns that could be associated with increased contact pressure and hyperadduction. Surprisingly, there is a dearth of studies that have investigated how the vibratory kinematics in individuals with phonotrauma differ from vocally healthy individuals. Some studies have investigated vibratory kinematics in phonotrauma but focused primarily on vibratory symmetry measures as opposed to measures associated with the closing phase dynamics (Chodara et al., 2012; Yamauchi et al., 2016). A single case study found high values of maximum vocal fold velocity in a person with phonotrauma (Lohscheller & Eysholdt, 2008). Another study investigated several measures in a group of children with vocal fold nodules, which showed increased closing velocity and increased speed quotient in children with nodules compared to healthy controls, but a similar study has not been conducted on adults with phonotrauma (Patel et al., 2016). Such information could elucidate how phonotraumatic lesions impact vocal fold vibratory kinematics and could provide valuable information to enhance assessment, treatment, and prevention of the disorders.

Objectives: The purpose of this study was to examine the differences in vocal fold vibratory kinematics, particularly those associated with the closing phase of the glottal cycle, between individuals with phonotrauma and vocally healthy controls. Our hypothesis was that vibratory closing phase dynamics will be increased in individuals with phonotrauma.

Methods: Twenty-six women with a diagnosis of phonotraumatic vocal fold lesions (14 with nodules, 12 with polyps) and 29 vocally healthy control participants underwent high-speed videoendoscopy via a transoral rigid scope while phonating on a sustained /i/ at a comfortable speaking pitch and volume in their modal register. Video exams were recorded at 4000 frames per second at 512 x 512 resolution and were collected during steady state production of the vowel. Exams were processed using the Glottis Analysis Tools software [version 2020] (Kist et al., 2021). Analyses were completed on 200 consecutive vibratory cycles. High-speed parameters (Closing Quotient, Maximum Area Declination Rate, Speed Index, and Amplitude-to-Length Ratio) were compared between patients and controls using independent t-tests and Cohen's d effect sizes.

Results: Patients with phonotrauma had significantly higher values of Amplitude-to-Length Ratio ($p = .004$; $d = .85$) indicating increased vibratory amplitude when normalized to glottal length, and higher Maximum Area Declination Rate ($p < .001$; $d = 1.27$) indicating higher vocal fold closing speed. There were no significant differences between participants with and without phonotrauma in Closing Quotient or

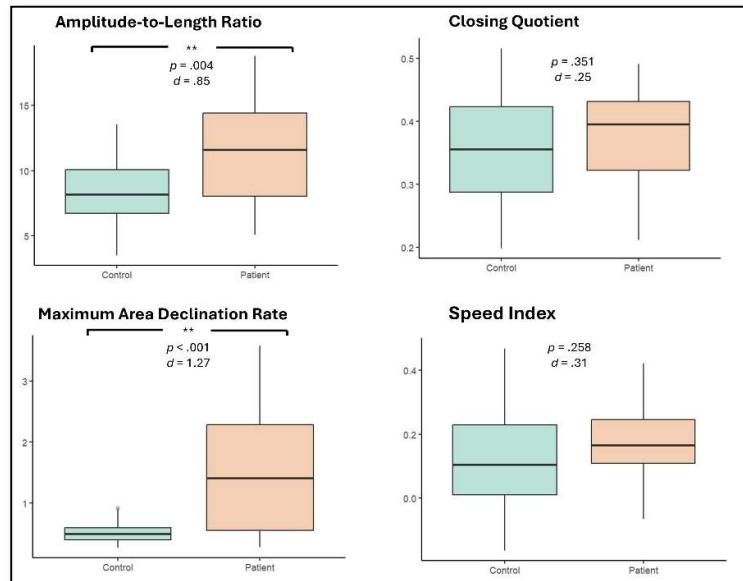


Figure 1. Boxplots of the four vibratory parameters comparing participants with phonotrauma (Patient, in pink) and vocally healthy controls (Control, in turquoise).

Speed Index. However, both parameters trended higher in the phonotrauma group with small effect sizes. Figure 1 shows boxplots for the distribution within each group.

Conclusions: Findings indicate that women with phonotrauma had increased vibratory amplitude and increased closing phase velocity compared to controls. These findings provide more evidence to suggest that individuals with phonotrauma have increased vocal fold impact stress. It is important to identify which outcome measures are specific to phonotraumatic voice use so that they can be investigated in longitudinal studies to disentangle whether the measures are causal to phonotrauma development or compensatory. This information can be used as a foundation to better discriminate between healthy and traumatic vocal fold vibratory kinematic patterns.

References:

- Chodara, A. M., Krausert, C. R., & Jiang, J. J. (2012). Kymographic characterization of vibration in human vocal folds with nodules and polyps. *The Laryngoscope*, 122(1), 58-65.
- 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.
- Kist, A. M., Gómez, P., Dubrovskiy, D., Schlegel, P., Kunduk, M., Echternach, M., Patel, R., Semmler, M., Bohr, C., & Dürr, S. (2021). A deep learning enhanced novel software tool for laryngeal dynamics analysis. *Journal of Speech, Language, and Hearing Research*, 64(6), 1889-1903.
- Lee, M., Mau, T., & Sulica, L. (2021). Patterns of Recurrence of Phonotraumatic Vocal Fold Lesions Suggest Distinct Mechanisms of Injury. *The Laryngoscope*.
- Lohscheller, J., & Eysholdt, U. (2008). Phonovibrogram visualization of entire vocal fold dynamics. *Laryngoscope*, 118(4), 753-758.
- Patel, R. R., Unnikrishnan, H., & Donohue, K. D. (2016). Effects of vocal fold nodules on glottal cycle measurements derived from high-speed videoendoscopy in children. *PLoS ONE*, 11(4), e0154586.
- Titze, I. R., Švec, J. G., & Popolo, P. S. (2003). Vocal dose measures: Quantifying accumulated vibration exposure in vocal fold tissues. *Journal of Speech, Language, and Hearing Research*, 46(4), 919-932.
- Yamauchi, A., Yokonishi, H., Imagawa, H., Sakakibara, K.-I., Nito, T., Tayama, N., & Yamasoba, T. (2016). Quantification of vocal fold vibration in various laryngeal disorders using high-speed digital imaging. *Journal of Voice*, 30(2), 205-214.

Comparison of Salivary Cortisol Levels Across Laryngeal Pathologies

Anumitha Venkatraman¹, Ruth Davis¹, Seth Pollak² Susan Thibeault¹

¹ Department of Surgery, Division of Otolaryngology Head and Neck Surgery, University of Wisconsin Madison, Madison, WI, USA

²Department of Psychology, University of Wisconsin Madison, Madison, WI, USA

Keywords: salivary cortisol, laryngeal pathologies, psychosocial stress

Abstract

Introduction & rationale: Psychosocial stress and laryngeal pathologies are inextricably and circularly linked.^{1,2} Although stress is thought to be an underlying factor in the development of primary Muscle Tension Dysphonia,³ the psychobiological framework to support this hypothesis is lacking.⁴ Prior literature has documented intrinsic and extrinsic laryngeal muscle activation patterns changes,^{4,5,6} or acoustic and aerodynamic measures following an acute stressor in healthy individuals.⁷ This study seeks to expand our biological knowledge of how psychosocial stress is linked to certain laryngeal pathologies.

In the presence of a psychosocial stressor, the Hypothalamic-pituitary-adrenal (HPA) axis produces cortisol which is transported throughout the body.⁸ HPA axis has a higher threshold of activation, activating in the presence of significant psychosocial stressors that occur for a longer duration (days to weeks or longer).⁹ Initially stressors result in HPA axis hyperactivity. However, when stressors occur repeatedly over a longer duration, these responses may be lessened or blunted (known as HPA axis hypoactivity).^{10,11} Up to 25 % of patients with voice problems report concurrent stress,¹² with significant impact on social and functional quality of life. Following appropriate referrals for voice problems, wait times to see available laryngologists may extend to months. Thus, there is a need to compare salivary baseline cortisol levels amongst patients with and without laryngeal pathologies.

Salivary cortisol is the most accurate method to measure free cortisol and is frequently used in literature on psychosocial stressors.^{13,14} Due to the natural diurnal rhythm of cortisol, baseline salivary cortisol should be collected in the same two-hour window across all patients before the intake of caffeine and food.¹⁵ Based on literature, salivary cortisol may also be affected by individualistic habits and experiences, including factors such as comorbid medical conditions, alcohol and marijuana use, and household income and composition.¹⁶⁻¹⁹ Due to limited literature,²⁰ we sought to account for these individualistic habits and experiences when comparing salivary cortisol between those with and without voice disorders.

Objectives: The aim of the study was to delineate changes in salivary cortisol amongst patients with and without different laryngeal pathologies, controlling for individualistic factors that affect cortisol (household income and composition, medical conditions and medications, alcohol and marijuana use). We hypothesized that baseline salivary cortisol levels in patients with laryngeal pathologies would be reduced compared to patients with non-voice related concerns.

Methods: New patients were recruited at UW Madison Otolaryngology Clinics by a laryngologist with the following inclusion criteria: patients > 18 years of age, presenting to the clinic in the morning, subsequently diagnosed with laryngeal pathology or a non-voice related concern. Exclusionary criteria included those who were presenting to the clinic for a follow up (i.e., had been to the clinic more than 2 visits), prior history of voice or upper airway problems, or had consumed caffeine or food two hours prior to the salivary cortisol collection. Following voice evaluation, patients met with a researcher, provided consent and completed a survey detailing education level, household income and composition, marijuana and alcohol intake, medical conditions, medications taken, sex and gender information on an online platform via UW Madison Qualtrics. They were asked to provide a salivary cortisol sample by spitting into a cryogenic vial with a coffee stirrer straw until 1 ml of saliva was collected. They were not provided water/candy to elicit saliva production prior to saliva collection. Cryogenic vials with saliva were stored at -20 degrees prior to evaluation. All salivary cortisol analysis was completed by UW Madison Assay Services using liquid chromatography-mass spectrometry (LC-MS). All salivary cortisol samples were taken in the same 2-hour time window. Preliminary data was analyzed with Kruskal Wallis tests for comparing salivary cortisol levels amongst those with and without voice disorder as well as individual factors separately due to limited sample size. Chi Squared tests were completed for categorical variables.

Results: Preliminary data indicates no significant differences in age across groups; Individuals diagnosed with a voice disorder (56.71 +/- 13 years, N=10, 6M, 4F, MTD = 7, Neurological voice problem = 3), and Control group (46.3 +/- 12 years, N=10, 7 M, 3 F, p = .13). Nineteen of the twenty participants were white Caucasian, with one participant identifying as Hispanic. Chi squared tests revealed no significant differences in education level, annual income, alcoholic drinks consumed in the last week, marijuana use, comorbid medical conditions, or specific medications taken between those who did and did not have a voice disorder (p > .05). **There were no significant differences in salivary cortisol amongst those with and without a voice disorder** (individuals with a voice disorder (.1702 + .08

ug/dL) and individuals without a voice problem (.2745 +/- .15 ug/dL, p = .041). There were similar salivary cortisol levels between types of voice disorders (MTD = .149 +/- 0.62 ug/dL, Control = .27 +/- .15 ug/dL, Neurological voice problems = .21 +/- .12 ug/dL). There were also no differences in salivary cortisol amongst sexes (Male, Female p = .394), age groups (18-30 years (n=2), 30-40 years (n=3), 50-60 years (n=5), 60-70 years (n=5), 70-80 years (n=5), p =.906) annual income (p = .702), or education level (p=.65), alcohol consumed in the last week (p =.69), or marijuana use (p=.75), type of comorbid medical conditions (p=.314) or those who do or do not primarily rely on their voice for their occupation (Yes (n=9)/No (n=11), p =.290).

Conclusions: Preliminary data indicates no differences in salivary cortisol levels in patients with and without laryngeal pathologies or across factors of demographic and lifestyle variables (age, sex, annual income, education, alcohol and marijuana use). A greater sample size – currently in process – may elucidate nuanced difference in salivary cortisol levels amongst patients with and without voice disorders.

References:

1. Nguyen-Feng, V. N., Asplund, A., Frazier, P. A., & Misono, S. (2021). Association between communicative participation and psychosocial factors in patients with voice disorders. *JAMA Otolaryngology—Head & Neck Surgery*, 147(3), 245-252.
2. Baker, J. (2008). The role of psychogenic and psychosocial factors in the development of functional voice disorders. *International journal of speech-language pathology*, 10(4), 210-230.
3. Roy, N., & Bless, D. M. (2000). Personality traits and psychological factors in voice pathology: a foundation for future research. *Journal of Speech, Language, and Hearing Research*, 43(3), 737-748.
4. Dietrich, M., & Verdolini Abbott, K. (2014). Psychobiological stress reactivity and personality in persons with high and low stressor-induced extralaryngeal reactivity. *Journal of Speech, Language, and Hearing Research*, 57(6), 2076-2089.
5. Dietrich, M., & Abbott, K. V. (2012). Vocal function in introverts and extraverts during a psychological stress reactivity protocol.
6. Perrine, B. L., & Scherer, R. C. (2020). Aerodynamic and acoustic voice measures before and after an acute public speaking stressor. *Journal of Speech, Language, and Hearing Research*, 63(10), 3311-3325.
7. Helou, L. (2014). *Intrinsic laryngeal muscle response to a speech preparation stressor: Personality and autonomic predictors* (Doctoral dissertation, University of Pittsburgh).
8. Perrine, B. L. (2018). *The Influence of stress on the voice* (Doctoral dissertation, Bowling Green State University).
9. Sapolsky, R.M., Romero, L.M., Munck, A.U., 2000. How do glucocorticoids influence stress responses? Integrating permissive, suppressive, stimulatory, and preparative actions. *Endocr. Rev.* 21, 55–89.
10. Essex, M.J., Shirtcliff, E.A., Burk, L.R., Ruttle, P.L., Klein, M.H., Slattery, M.J., Kalin, N. H., Armstrong, J.M., 2011. Influence of early life stress on later hypothalamic-pituitary-adrenal axis functioning and its covariation with mental health symptoms: a study of the allostatic process from childhood into adolescence. *Dev. Psychopathol.* 23, 1039–1058.
11. Shirtcliff, E. A., Hanson, J. L., Phan, J. M., Ruttle, P. L., & Pollak, S. D. (2021). Hyper-and hypo-cortisol functioning in post-institutionalized adolescents: The role of severity of neglect and context. *Psychoneuroendocrinology*, 124, 105067.
12. Dietrich, M., Abbott, K. V., Gartner-Schmidt, J., & Rosen, C. A. (2008). The frequency of perceived stress, anxiety, and depression in patients with common pathologies affecting voice. *Journal of voice*, 22(4), 472-488.
13. Hellhammer, D. H., Wüst, S., & Kudielka, B. M. (2009). Salivary cortisol as a biomarker in stress research. *Psychoneuroendocrinology*, 34(2), 163-171.
14. Pan, X., Wang, Z., Wu, X., Wen, S. W., & Liu, A. (2018). Salivary cortisol in post-traumatic stress disorder: a systematic review and meta-analysis. *BMC psychiatry*, 18, 1-10.
15. Hansen, A. M., Garde, A. H., & Persson, R. (2008). Sources of biological and methodological variation in salivary cortisol and their impact on measurement among healthy adults: a review. *Scandinavian journal of clinical and laboratory investigation*, 68(6), 448-458.
16. Hajati, A., Diez-Roux, A., Franklin, T. G., Seeman, T., Shrager, S., Ranjit, N., ... & Kirschbaum, C. (2010). Socioeconomic and race/ethnic differences in daily salivary cortisol profiles: the multi-ethnic study of atherosclerosis. *Psychoneuroendocrinology*, 35(6), 932-943.
17. Dowd, J. B., Ranjit, N., Do, D. P., Young, E. A., House, J. S., & Kaplan, G. A. (2011). Education and levels of salivary cortisol over the day in US adults. *Annals of Behavioral Medicine*, 41(1), 13-20.
18. Simon, S. G., Jammer, L. D., Dent, A. L., Granger, D. A., & Riis, J. L. (2023). Hypothalamic-pituitary-adrenal and sympathetic nervous system responses to social evaluative stress in chronic cannabis users and non-users. *Addictive Behaviors*, 136, 107489.
19. King, A., Munisamy, G., de Wit, H., & Lin, S. (2006). Attenuated cortisol response to alcohol in heavy social drinkers. *International journal of psychophysiology*, 59(3), 203-209.
20. Pisanski, K., Nowak, J., & Sorokowski, P. (2016). Individual differences in cortisol stress response predict increases in voice pitch during exam stress. *Physiology & behavior*, 163, 234-238.



Patients presenting to
UW Madison
Otolaryngology clinic:

- For the first or second time
- Presenting in the morning
- Had not eaten or drank 2 hours prior to meeting with the researcher

Recruited by a
laryngologist following
diagnosis with:

- Muscle Tension Dysphonia
- Neurological Voice problem
- Non-voice related concern

Completed a survey
and provided a salivary
cortisol sample. All
participants provided
samples in the same 2
hour window.

Figure 1: Experimental Protocol

Targeted Nebulized ECM Therapy for Vocal Fold Healing

Nour Awad, MD¹; Alexander D. Karabachev, MD¹; Ramair Colmon²; Camilo A Mora-Navarro, Ph.D³; Donald O. Freytes, Ph.D²; Gregory R. Dion, MD, MS, FACS¹

¹University of Cincinnati College of Medicine, Department of Otolaryngology – Head and Neck Surgery

²The Joint Department of Biomedical Engineering, University of North Carolina at Chapel Hill and North Carolina State University

³Department of Chemical Engineering, University of Puerto Rico-Mayaguez

Keywords: Vocal fold scar, Vocal fold lamina propria (VFLP), Nebulization therapy

Abstract

Introduction & rationale: Voice disorders represent the most prevalent communication disorder in the United States [1], affecting approximately 1 in 8 adults annually and leading to significant financial, occupational, and psychosocial burdens, including depression and anxiety [2, 3, 4, 5].

Vocal fold (VF) scarring, a common cause of voice disorders, arises from various factors such as neck trauma, radiation therapy, chronic gastroesophageal reflux, upper respiratory tract infections, smoking, and aging [4, 7]. It is characterized by fibrotic tissue replacing healthy VF mucosa, impairing vibration and voice quality. Current therapeutic strategies for VF scarring often involve tissue engineering approaches, particularly the injection of biomaterials aimed at restoring function or promoting tissue regeneration [8]. Among these, extracellular matrix (ECM) scaffolds derived from decellularized porcine tissues are highly promising due to their low immunogenicity, inherent bioactivity, antibacterial properties (via cryptic peptides), and ability to facilitate constructive tissue remodeling while minimizing scar formation [9]. ECM-derived hydrogels, fabricated through decellularization, powdering, lyophilization, and pepsin digestion, are widely utilized in regenerative medicine for their capacity to enhance tissue remodeling and reduce fibrosis. However, many ECM-based treatments require invasive endoscopic procedures or repeated scaffold injections, necessitating clinical supervision and multiple follow-up visits, thereby increasing patient burden. A less invasive approach for ECM delivery, such as nebulized ECM hydrogels, could provide an effective and patient-friendly alternative for VF tissue regeneration, minimizing procedural risks while enhancing therapeutic outcomes.

The current study aims to evaluate the use of a nebulized vocal fold lamina propria ECM hydrogel (VFLP-ECM) as a novel non-invasive regenerative therapy for vocal fold injury. The nebulized ECM hydrogel was designed to enable targeted upper airway deposition, allowing for localized tissue regeneration while avoiding the need for surgical implantation.

Objectives: This study aims to evaluate the antifibrotic and antibacterial properties of aerosolized laryngeal-derived extracellular matrix (ECM), initially formulated as an injectable, administered via a nebulization approach.

Methods: A porcine-derived, nebulized vocal fold lamina propria extracellular matrix (VFLP-ECM) analog was developed for targeted vocal fold delivery. Bilateral vocal fold injuries were induced endoscopically in Sprague Dawley rats using microforceps. To minimize laryngospasm and vocal fold movement, 100 µL of 2% topical lidocaine was applied. A custom 3D-printed laryngoscope enabled instrument access and visualization via a surgical microscope. Rats were positioned dorsally on a platform tilted 30° for optimal exposure.

Forty-eight hours post-injury, animals were randomly assigned to one of four treatment groups: (1) no-treatment control, (2) VFLP-ECM, (3) phosphate-buffered saline (PBS), or (4) collagen type I (COL), with six rats per group.

Fourteen days after injury, the larynges were harvested for histological analysis using Hematoxylin and Eosin and Masson's trichrome staining. Tissue characteristics, including collagen density, lamina propria thickness, and cellular infiltration, were evaluated to assess the regenerative effects of nebulized ECM treatment.

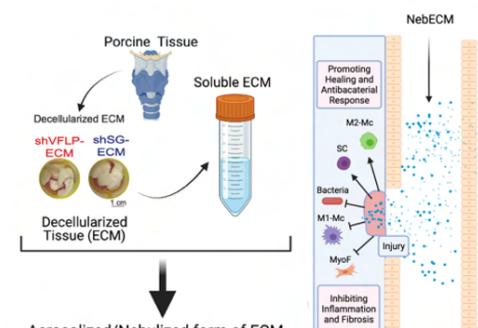


Figure 1. Solubilized ECM for nebulization procedure

Results: Results from histological examination revealed differences in vocal fold tissue characteristics following nebulized extracellular matrix (ECM) treatment. The ECM-treated group exhibited a notable reduction in dense collagen deposition compared to both PBS-treated and untreated control groups, suggesting potential improvements in tissue elasticity. Furthermore, an increase in lamina propria thickness was observed in the ECM-treated specimens relative to the control group, with this enhanced tissue layer depth consistently noted across multiple samples. Cellular analysis demonstrated a higher total cell count within the lamina propria of the ECM-treated group compared to the PBS-treated cohort.

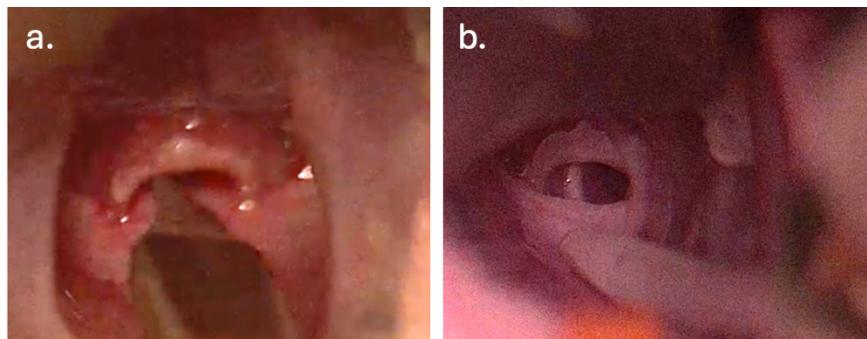


Figure 2. Endoscopic images of rat vocal fold, showing a) mechanism of injury induced by micro-forceps, b) nebulization procedure. Showing insertion of the tube into the vocal fold

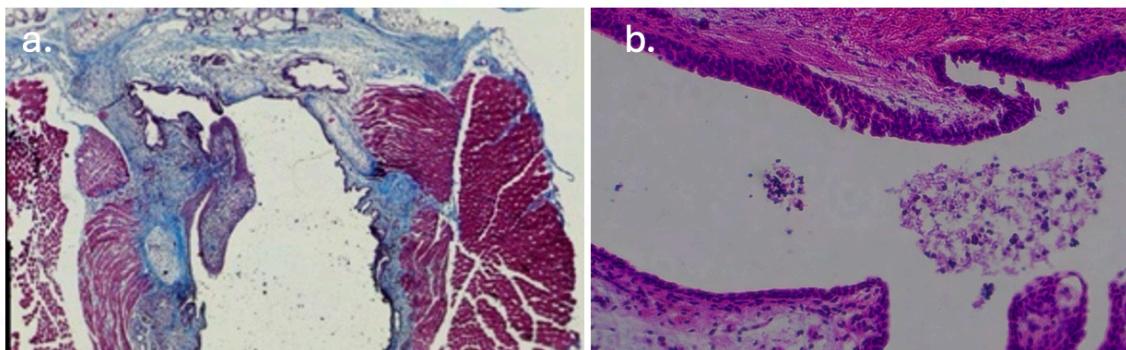


Figure 3. a) Trichrome stain of vocal fold histology at day 14 showing injury, b) vocal fold histology after nebulization

Conclusion: These findings suggest that nebulized delivery of vocal fold lamina propria extracellular matrix enhances vocal fold tissue characteristics and may support tissue regeneration during acute wound healing. The study demonstrates a potential application for nebulized decellularized ECM in vocal fold regenerative medicine, offering a less invasive alternative to current treatment methods.

References:

1. Ramig, L. O., & Verdolini, K. (1998). Treatment efficacy: Voice disorders. *Journal of Speech, Language, and Hearing Research*, 41(S101–S116).
2. Roy, N., Merrill, R. M., Gray, S. D., & Smith, E. M. (2005). Voice disorders in the general population: prevalence, risk factors, and occupational impact. *The Laryngoscope*, 115(11), 1988–1995. <https://doi.org/10.1097/01.mlg.0000179174.32345.41>
3. Cohen, S. M., Kim, J., Roy, N., Asche, C., & Courey, M. (2012). Direct health care costs of laryngeal diseases and disorders. *The Laryngoscope*, 122(7), 1582–1588. <https://doi.org/10.1002/lary.23189>
4. Benninger, M. S., Alessi, D., Archer, S., Bastian, R., Ford, C., Koufman, J., Sataloff, R. T., Spiegel, J. R., & Woo, P. (1996). Vocal fold scarring: current concepts and management. *Otolaryngology--head and neck surgery : official journal of American Academy of Otolaryngology-Head and Neck Surgery*, 115(5), 474–482. <https://doi.org/10.1177/019459989611500521>
5. Hansen, J. K., & Thibault, S. L. (2006). Current understanding and review of the literature: vocal fold scarring. *Journal of voice : official journal of the Voice Foundation*, 20(1), 110–120. <https://doi.org/10.1016/j.jvoice.2004.12.005>
6. Hansen, J. K., & Thibault, S. L. (2006). Current understanding and review of the literature: vocal fold scarring. *Journal of voice : official journal of the Voice Foundation*, 20(1), 110–120. <https://doi.org/10.1016/j.jvoice.2004.12.005>
7. Gracioso Martins, A. M., Biehl, A., Sze, D., & Freytes, D. O. (2022). Bioreactors for Vocal Fold Tissue Engineering. *Tissue engineering. Part B, Reviews*, 28(1), 182–205. <https://doi.org/10.1089/ten.TEB.2020.0285>
8. Badylak, S. F., Freytes, D. O., & Gilbert, T. W. (2009). Extracellular matrix as a biological scaffold material: Structure and function. *Acta biomaterialia*, 5(1), 1–13. <https://doi.org/10.1016/j.actbio.2008.09.013>

Discovering the Effects of Vascular Lesions of the Vocal Fold via Computational Modeling

Manoela Neves¹, Rana Zakerzadeh¹

¹ Department of Biomedical Engineering, Duquesne University, Pittsburgh, PA, U.S.A.

Keywords: computational vocal fold models, fluid-structure interaction (FSI), vascular lesions

Abstract

Introduction & rationale: Vocal folds (VFs) are complex, multilayered structures whose biomechanical and aerodynamic properties are crucial for phonation. Understanding the impact of structural variations and pathological changes on VF dynamics can significantly aid in clinical treatments for voice disorders. This study investigates the effects of VF multilayer structural composition and the presence of vascular lesions on phonatory behavior, using computational fluid-structure interaction (FSI) models to simulate the influence of these lesions on glottal airflow and tissue oscillation. VF vascular lesions form due to excessive vibration exposure and the resulting vibration-induced rises in capillary pressure, causing the enlargement or rupture of blood vessels along the VF mucosa layer, also known as the cover. Consequently, VF hemorrhage occurs when one or more of the enlarged vessels breaks open, and the extravasated blood diffuses through Reinke's space. Even a small amount of blood weighs down the VF tissue. The presence of VF vascular lesions leads to vocal fatigue and dysphonia due to increased mass and localized changes in tissue viscoelastic properties, disrupting VF vibratory symmetry and altering airflow resistance through the glottis. Despite their clinical relevance, the integration of these lesions in computational simulations of VFs remains unexplored. This study aims to fill this gap by exploring how vascular lesions affect VF vibratory behavior and glottal airflow dynamics.

Objectives: The objective is to discover vascular lesions' impact on vibratory behaviors of VF tissue and the resulting vocal mechanisms. The impacts of VF vascular lesions on phonation are analyzed via the biomechanical and aerodynamic properties of a multilayered subject-specific VF model which includes both healthy and lesion-affected tissues. Specifically, this study aims to assess how two common pathological conditions; a blood clot polyp and a dilated blood vessel (stationary crack filled with fluid) embedded in the VF cover layer, disrupt VF oscillations and alter airflow resistance. By incorporating anatomically accurate larynx geometry and anisotropic material properties, the presented approach will advance our understanding of VF dynamics and provide insights into pathological conditions, exploring the potential for early disease detection and therapeutic interventions.

Methods: A three-dimensional, bilayer VF model is constructed based on STL files derived from a human larynx CT scan, which considers the variation in mechanical properties of cover and body layers. VFs designed to fit into the laryngeal space and symmetry along the sagittal line is assumed. The glottal airflow is considered laminar and modeled by the unsteady, viscous, incompressible Navier-Stokes equations, while the momentum equation for the balance of total forces is used to describe the VFs and lesions deformations. Physiologically realistic subglottal pressure at the inlet and supraglottal gauge pressure at the outlet are applied as boundary conditions to replicate human phonation conditions. Fluid-structure interaction (FSI) approach is adopted to simulate the coupling of aerodynamics and tissue dynamics sub-problems in ANSYS Workbench platform (Zakerzadeh et al., 2025). Furthermore, soft tissues including vascular lesions and VFs are modelled as a fluid-saturated porous medium (McCollum et al., 2023).

Results & conclusions: Two FSI models, one with lesions and one without, are created to assess the effects of lesion presence. Analysis of glottal airflow dynamics and tissue vibrational patterns between these two models is used to determine the impact of vascular lesions on biomechanical characteristics of phonation. Additionally, VF intravascular pressure during various phonation conditions such as loud speaking and singing is explored. The results underscore the importance of incorporating layered VF structures in phonation models, highlighting the significant role of the cover layer in facilitating efficient phonation and the disruptive effects of lesions on VF dynamics. An assessment of the VF vibration between healthy and abnormal tissue revealed distinct differences in frequency and amplitude. The symmetric VF motion is associated with a healthy tissue and the asymmetric one to the case when lesion is present. Introducing lesions to the cover layer also disrupted glottal closure and airflow patterns, resulting in irregular vibrations and longer phonation cycles.

Acknowledgments: The authors acknowledge support from NSF under grant CMMI 2338676.

References:

- McCollum, I., Badr, D., Throop, A., & Zakerzadeh, R. (2023). Biotransport in human phonation: Porous vocal fold tissue and fluid-structure interaction. *Physics of Fluids*, 35(12).
- Zakerzadeh, R., McCollum, I., & Neves, M. (2025). Coupled processes of tissue oxygenation and fluid flow in biphasic vocal folds. *International Journal of Heat and Mass Transfer*, 238, 126494.

A Platform for Monitoring Ecological Vocal Efficiency

Ahmed M. Yousef¹, Emma C. Willis^{1,2}, Vahni Tagirisa¹, Robert E. Hillman¹⁻³, Daryush D. Mehta¹⁻³

¹ Center for Laryngeal Surgery and Voice Rehabilitation, Massachusetts General Hospital, Boston, MA, USA

² School of Health & Rehabilitation Sciences, MGH Institute of Health Professions, Boston, MA, USA

³ Harvard Medical School, Boston, MA, USA

Keywords: Ambulatory Voice Monitoring; Vocal Efficiency; Voice Disorders

Abstract

Introduction & rationale: Voice disorders often present as inefficient vocal function and behavior during speech, which can significantly impact effective communication in social and professional interactions [1]. Vocal efficiency (VE) is an acoustic-aerodynamic measure representing the efficiency of the larynx to convert aerodynamic energy into acoustic energy. Traditionally, VE is defined by the ratio of the radiated acoustic power (computed from sound pressure level (SPL)) to aerodynamic power (a combination of subglottal pressure (Ps) and glottal airflow) during voice production [2]. Due to the variability in airflow, other simplified formulations of this ratio were adapted in the literature to quantify VE such as a ratio of SPL to Ps [3-5]. The simplified ratio SPL/Ps has demonstrated clinical validity to differentiate between patients with voice disorders from vocally healthy individuals [5] as well as sensitivity to pre- and post-treatment effects following surgical procedures and/or voice therapy [3,4].

Objectives: Although VE measures provide valuable insights into vocal function and have great potential for voice assessment, such measures have traditionally been limited to controlled laboratory environments or brief evaluations during clinical visits. In this ongoing work, our goal is to measure the vocal efficiency ratio in real-world settings by capturing data during daily activities and routines. We present a platform designed to monitor ecological vocal efficiency (EVE), enabling the quantification of meaningful changes in vocal function and behavior within natural, everyday environments. The EVE index utilizes the simplified SPL/Ps ratio. We aim to investigate the performance of the ambulatory EVE index to discriminate patients with voice disorders from vocally healthy controls. Our central hypothesis is that the EVE index can accurately quantify differences in vocal efficiency between individuals with voice disorders and vocally healthy controls, offering clinically meaningful insights into how patients use their voice outside of traditional clinical and laboratory settings. It is also hypothesized that the mean and variability of the EVE index in patients would be lower, reflecting reduced vocal efficiency and a narrower range of vocal behavior associated with voice disorders.

Methods: Our ongoing study enrolls individuals with voice disorders characterized by inefficient vocal function in the lab or clinic: vocal fold lesions, muscle tension dysphonia, and unilateral vocal fold paralysis. In-lab data collection and diagnoses are performed at the Massachusetts General Hospital Voice Center. Each patient is paired with a vocally healthy individual matched by sex, age, and occupational vocal demands. Participants complete a laboratory session to establish baseline vocal efficiency values. To do this, participants are instructed to produce /p/-vowel sequences (/pa/, /pi/, /pu/) at varying loudness (loud to soft) and pitch (comfortable, high, low) levels. Simultaneous recordings are obtained of the microphone, oral airflow, intraoral pressure, electroglottograph, and accelerometer signals. The microphone and accelerometer sensors are calibrated for accurate estimation of voice SPL and Ps measures for field measurements.

Following the in-lab session, all participants wear a smartphone-based voice monitoring platform for three waking days to track the EVE index and ecological momentary assessment of vocal effort during their daily activities. We use a state-of-the-art dual-sensor, wireless voice monitor capable of measuring vocal SPL with a microphone and estimating Ps using a neck-surface accelerometer [1,6,7]. By combining these measurements, we compute the EVE index (SPL/Ps), offering a new approach to assessing vocal function in natural, real-world environments. Building on prior research, we have developed advanced methods to accurately predict Ps from accelerometer-based neck vibration data [8,9]. We then analyze ambulatory EVE indices, including statistical metrics such as mean, standard deviation, skewness, kurtosis, and 5th and 95th percentiles, to determine differences between each patient group and their matched controls using independent-sample t-tests.

Results: Laboratory assessments of the EVE index effectively distinguished patients with voice disorders from their matched control groups, demonstrating its diagnostic utility. These findings lay the groundwork for extending the application of EVE monitoring to the natural environment. Building on these laboratory findings, our preliminary in-field results highlight the feasibility and potential clinical value of monitoring the ambulatory EVE index in real-world settings.

Figure 1 shows an example of ambulatory data collected from a female with vocal fold nodules, who wore the voice monitor device for 7 hours, and a vocally healthy matched control for comparison. These data demonstrate the feasibility of tracking the percent phonation time and the corresponding EVE index throughout the day, represented by maximum and median values, for each subject. The distribution histograms are also presented, for SPL (dB SPL), Ps (cm H₂O), and EVE index (dB/dB) for the entire 7-hour period of monitoring. The patient's EVE index distribution shows both a lower average and reduced variability compared to the vocally healthy female, reflecting decreased vocal efficiency and a more constrained range of EVE values. This observation is further confirmed by the patient's reduced average EVE

index, which falls below 4 dB/dB, compared to around 5.5 dB/dB in the healthy control subject. These example results demonstrate the potential for the EVE index to offer clinicians a promising measure of vocal function, capturing how patients use their voice in daily life beyond the clinical setting. Comparisons between patients with different types of voice disorders and vocally healthy matched controls will be presented.

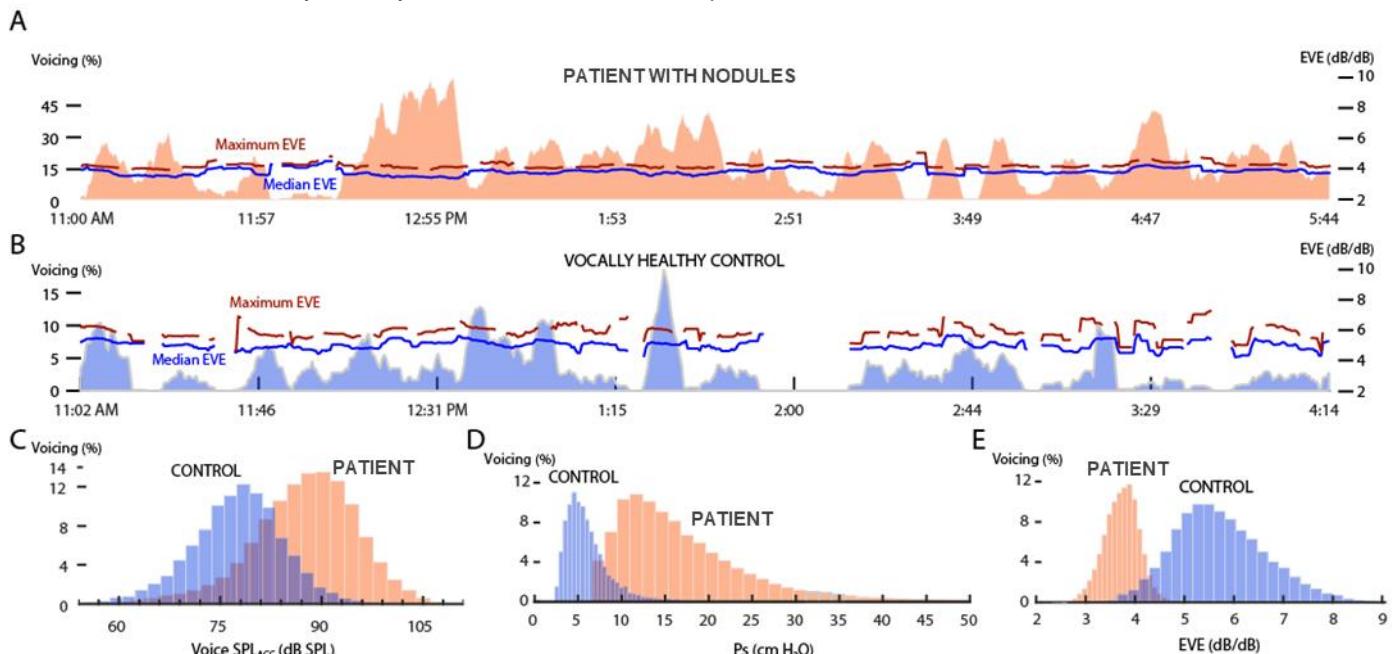


Figure 1. Feasibility study of obtaining the ecological vocal efficiency (EVE) index along with percent phonation during the day for a (A) female patient with vocal fold nodules and a (B) vocally healthy female subject. (C) The patient's higher voice sound pressure level (SPL) is generated by (D) a disproportionately high subglottal pressure (Ps), leading to a (E) reduction in vocal efficiency as measured by the distribution of the EVE index.

Conclusions: This study aims to develop and validate the EVE index as a reliable metric for assessing vocal function in real-world environments. We investigated the capability of the EVE index to differentiate between patients with voice disorders and vocally healthy matched controls using a wearable ambulatory voice monitoring device that integrates accelerometer and microphone signals. Based on our preliminary findings, the EVE index effectively differentiates patients with voice disorders from healthy individuals. The EVE index provides clinicians with valuable insights into patients' vocal behavior in natural environments, overcoming the limitations of in-clinic assessments. It also holds potential as an early-warning indicator to help prevent the development/progression of voice disorders, a biofeedback tool for voice therapy, and a clinical outcome measure for laryngeal surgery and voice therapy.

References:

- [1] Van Stan, J. H., Maffei, M., Masson, M. L. V., Mehta, D. D., Burns, J. A., and Hillman, R. E. (2017). Self ratings of vocal status in daily life: Reliability and validity for patients with vocal hyperfunction and a normative group. *American Journal of Speech-Language Pathology*, 26(4):1167-1177. PMCID: PMC5945061.
- [2] Titze, I. R. (1992). Vocal efficiency. *Journal of Voice*, 6(2):135-138.
- [3] Zeitels, S. M., Hillman, R. E., Desloge, R. B., and Bunting, G. A. (1999). Cricothyroid subluxation: A new innovation for enhancing the voice with laryngoplastic phonosurgery. *Annals of Otology Rhinology and Laryngology*, 108(12):1126-31.
- [4] Zeitels, S. M., Hillman, R. E., Franco, R. A., and Bunting, G. W. (2002). Voice and treatment outcome from phonosurgical management of early glottic cancer. *Annals of Otology, Rhinology, and Laryngology*, 111 (Supplement 190)(12 Part 2):1-20.
- [5] Espinoza, V. M., Zañartu, M., Van Stan, J. H., Mehta, D. D., and Hillman, R. E. (2017). Glottal aerodynamic measures in women with phonotraumatic and nonphonotraumatic vocal hyperfunction. *Journal of Speech, Language, and Hearing Research*, 60(8):2159-2169. PMCID: PMC5829799
- [6] Mehta, D. D., Zañartu, M., Feng, S. W., Cheyne II, H. A., and Hillman, R. E. (2012). Mobile voice health monitoring using a wearable accelerometer sensor and a smartphone platform. *IEEE Transactions on Biomedical Engineering*, 59(11):3090-3096. PMCID: PMC3539821.
- [7] Mehta, D. D., Van Stan, J. H., Zañartu, M., Ghassemi, M., Guttag, J. V., Espinoza, V. M., Cortés, J. P., Cheyne II, H. A., and Hillman, R. E. (2015). Using ambulatory voice monitoring to investigate common voice disorders: Research update. *Frontiers in Bioengineering and Biotechnology*, 3(155):1-14. PMCID: PMC4607864.
- [8] Lin, J. Z., Espinoza, V. M., Marks, K. L., Zañartu, M., and Mehta, D. D. (2020). Improved subglottal pressure estimation from neck-surface vibration in healthy speakers producing non-modal phonation. *IEEE Journal of Selected Topics in Signal Processing*, 14(2):449-460.
- [9] Marks, K. L., Lin, J. Z., Burns, J. A., Hron, T. A., Hillman, R. E., and Mehta, D. D. (2020). Estimation of subglottal pressure from neck surface vibration in patients with voice disorders. *Journal of Speech, Language, and Hearing Research*, 63(7):2202-2218.

Dynamic 3D Vocal Fold MRI for Quantification of Vocal Fold Oscillations in Different Voice Production Mechanisms

Louisa Traser^{1,2}, Johannes Fischer^{2,3}, Paula Jordan^{2,3}, Fiona Stritt^{1,2}, Marie Körberlein⁴, Jonas Kirsch⁴, Stefanie Rummel⁵, Bernhard Richter^{2,3}, Michael Bock^{2,3}, Matthias Echternach⁴

¹Freiburg Institute for Musicians' Medicine, Medical Center, University of Freiburg, Freiburg, Germany

²Faculty of Medicine, University of Freiburg, Freiburg, Germany

³Dept. of Radiology, Medical Physics, Medical Center University of Freiburg, Faculty of Medicine, University of Freiburg, Freiburg, Germany

⁴Department of Otorhinolaryngology, Ludwig-Maximilians-Universität München, Division of Phoniatrics and Pediatric Audiology, LMU Klinikum, Munich, Germany

⁵Institut Rummel, Frankfurt, Germany

Keywords: Magnetic Resonance Imaging, Vocal folds, High-Speed Imaging

Abstract

Introduction & Rationale: Voice research is using endoscopic views on the vocal folds from above (VF) to observe laryngeal adjustments in a 2D plane. However, recent computer simulations (1–3) and ex-vivo or canine studies (4–6) underscore the importance of a three-dimensional assessment of VF vibration. Specifically, the VF vertical thickness and the medial surface shape play a significant role in shaping glottal closure patterns. Here, increasing vertical thickness has been associated with prolonged closure duration within a vibratory cycle due to the enhanced phase delay between the lower and upper margins (5). Additionally, thicker VFs are believed to counteract aerodynamic forces, promoting glottal closure under higher subglottal pressures (3). These geometrical features can be measured non-invasively using dedicated MRI methods for VF imaging (7,8).

Objectives: To measure medial surface configuration, vertical and horizontal VF displacement and vertical thickness during phonation *in vivo* for clinical and pedagogical applications, to gain insights into the mechanisms of voice disorders and to address specific demands in vocal training.

Methods: This study presents initial findings from 3D MRI measurements of VF motion during phonation using various voice production mechanisms, performed by a professionally trained singer (co-author SR). The "figures" framework of Estill Voice Training® was used to describe the voice production mechanisms (9). Here, "Thick" and "Thin" represent the intended thickness of the vocal folds, while "Stiff" refers to a mode with abduction in the posterior part of the VF leading to breathy phonation. "Thick & False Vocal Folds (FVF) retract" and "Thick & FVF constrict" combines thick VF with the ab-/adduction of the FVF. "Thick & aryepiglottic sphincter (AES)" involves additional activation of the AES with thick vocal folds and the FVF in a neutral position.

MRI data were acquired on a clinical 3T MRI scanner (Siemens PrismaFit) using a zero echo time (ZTE) pulse sequences introduced by (8). A circular radiofrequency coil was positioned at the larynx to sensitively acquire the MR signal, achieving an image resolution of 0.8 mm over a scan time of 4 min 20 s. During the MRI acquisition, the singer phonated at a constant fundamental frequency ($f_0 = 200$ Hz) and paused intermittently for breathing. Laryngeal positional shifts between phonations were corrected using a navigator acquisition. Positional corrections were applied retrospectively using the Fourier shift theorem.

To reconstruct VF oscillation, MR data were binned into 10 time frames based on VF motion phase during signal acquisition. Acoustic signals were recorded via a microphone placed at the

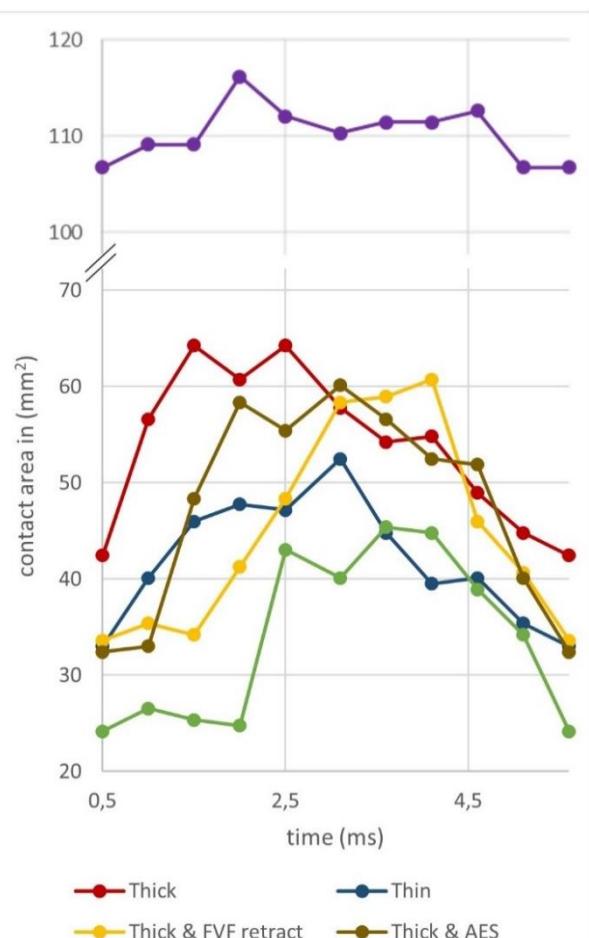


Figure 1: Vocal fold contact area in 10 frames of an averaged VF vibration for 6 voice production mechanisms

The 16th International Conference on Advances in Quantitative Laryngology, Voice, and Speech Research volunteer's mouth and processed with a low-pass Butterworth filter to minimize MRI noise interference and extract the f_0 . Data for one VF vibration were sorted into 10 distinct phases of the oscillatory cycle. In the resulting dynamic image sequence, VF contours were segmented in each frame, and geometric shape parameters were quantified using threshold-based pixel analysis. Based on this, the maximum VF displacement in horizontal and vertical dimensions during phonation was determined in the coronal plane. Furthermore, the vocal fold contact area was measured in the midsagittal plane. As a proof of principle, in a supplementary session, transnasal endoscopy of the larynx was performed in the same subject, phonating the same voice production mechanisms using high-speed videoendoscopy (HSV) at 20,000 fps with a spatial resolution of 386x320 pixels as introduced by (9). Acoustic pressure, electroglottographic (EGG) signals, and calibrated sound pressure levels (SPL) were recorded simultaneously. The glottal area waveform (GAW) according to (10) was segmented for further analysis.

Results: For different voice production types the VF contact area varied: "Stiff" shows the smallest maximum contact area, while "Thick & FVF constrict" exhibits the largest. The VF contact curves vary in the position of their maximum and their variance. A leftward skew is observed for "Thick", the maximum lies centrally for "Thin", and shifts to the right for "Stiff". "Thick & FVF constrict" shows remarkably small variance in VF contact area compared to other phonation types, which is also reflected in its minimal VF displacement, particularly in the vertical direction. The highest VF displacement occurs in "Thick" (horizontal) and "Thick and FVF retract"/"Stiff" (vertical). These motion parameters align with the open quotient (OQ) values of the GAW and EGG: the smallest opening is seen in "Thick & FVF constrict", followed by "Thick & AES", "Thick", "Thick & FVF retract", "Thin" and largest in "Stiff".

Discussion: The parameters quantified with 3D VF MRI align well with the established OQ derived from EGG and GAW. The VF contact area curve similarly reflects properties of the relative contact area represented by the EGG curve. Notably, a skewing to the left was observed with increasing contact area and intended vertical VF thickness, consistent with previous findings (11). Unlike EGG, however, MRI-based parameters additionally capture absolute differences in contact area. In the presented data, the adduction or abduction of the FVF not only led to an decrease or increase in OQ, but also resulted in significant changes in VF displacement, particularly in the vertical direction. This is corroborated by spatial information from MR images, which reveal that FVF abduction is associated with an expansion of the supralaryngeal space, while FVF adduction causes narrowing through closer proximity to the true vocal folds. As previously suggested by simulations (3), *in vivo* VF thickness appears to be directly correlated with contact area. Quantification of VF displacement demonstrated stronger vertical than horizontal motion, in line with findings from laser-based 3D high-speed imaging (12). The data further suggests that this relationship depends on the voice production mechanism. Since endoscopy is limited in capturing vertical VF motion, these findings highlight the critical need to evaluate such parameters to advance the understanding of vocal fold vibration, including its dysfunctions.

Conclusions: This first application of dynamic 3D VF MRI in singers shows that MRI-derived parameters correlate well with established values, while the additional insights gained through quantification are substantial. Future efforts should focus on refining analysis methods, including 3D reconstruction of motion and the automated extraction of key parameters characterizing VF vibration, such as convergence and divergence angles, overall thickness, and 3D displacement. Additionally, measurements in patients with voice disorders are planned to further expand the clinical and scientific utility of this technique.

Acknowledgments: Grant support from the Deutsche Forschungsgemeinschaft (DFG) under grant numbers FI 2803/1-1 and TR 1491/4-1 is gratefully acknowledged

References:

- Lehoux S et al. A Methodology to Quantify the Effective Vertical Thickness of Prephonatory Vocal Fold Medial Surface. J Voice. 2024
- Alipour F et al. Vocal fold bulging effects on phonation using a biophysical computer model. J Voice. 2000
- Zhang Z. Vocal Fold Vertical Thickness in Human Voice Production and Control: A Review. Journal of Voice; 2023.
- Berry DA et al. High-speed digital imaging of the medial surface of the vocal folds. J Acoust Soc Am. 2001
- Zhang Z et al. Effect of changes in medial surface shape on voice production in excised human larynges. J Acoust Soc Am; 2019.
- Khosla S et al. Direct simultaneous measurement of intraglottal geometry and velocity fields in excised larynges. Laryngoscope. 2014
- Fischer J et al. Sub-millisecond 2D MRI of the vocal fold oscillation using single-point imaging with rapid encoding. Magn Reson Mater Physics, Biol Med. 2022
- Fischer JF et al. Isotropic 3D Sub-millimeter MRI of the Vocal Fold Oscillation with Sub-millisecond Temporal Resolution. In: ISMRM Annual Meeting Proceedings. 2024.
- Steinhauer K et al. The Estill Voice Model: Theory and Translation. 2017.
- Echternach M et al. Laryngeal evidence for the first and second passaggio in professionally trained sopranos. PLoS One. 2017
- Titze IR. Interpretation of the electroglottographic signal. J Voice 1990
- Semmler M et al. Clinical relevance of endoscopic three-dimensional imaging for quantitative assessment of phonation. Laryngoscope. 2018

Task	OQ GAW	OQ EGG	vertical Displ.	horizontal Displ.
Thick	0,34	0,39	2,30	3,1
Thin	0,59	0,50	2,30	0,8
Stiff	0,86	0,70	3,10	2,3
Thick & AES	0,32	0,31	2,30	1,5
Thick & FVF retract	0,49	0,38	3,10	2,3
Thick & FVF constrict	0,23	0,25	0,80	1,5

Table 1: Open Quotient derived from GAW and EGG, maximal vertical and horizontal vocal fold displacement (in mm) derived from MRI data in 6 voice production mechanisms.

Electroglottographic Analysis of the Voice in Mandarin-Speaking Children

Wang Binghao, Lu Yao, Liang Changwei, Wu Xiyu, Kong Jiangping

Laboratory of Language Sciences and Department of Chinese Language and Literature,
Peking University, Beijing, China

Keywords: Children's voice; Electroglottography; Vocal fold vibration

Abstract

Introduction & rationale: Fundamental frequency (F0), open quotient (OQ) and speed quotient (SQ) are three features of vocal fold vibration indicating distinct phonation types [1]. F0, a time domain feature, refers to the frequency in the quasi-periodic vibration of vocal cords; OQ, a frequency domain feature, represents the ratio of the vocal fold open phase duration to the total vibratory period; and SQ, another frequency domain feature, is defined as the ratio of the vocal fold opening phase duration to the closing phase duration [2]. Technological advancements in voice analysis have allowed the extraction of the above-mentioned parameters by Electroglottography. This method has been adopted to analyse vocal fold vibration in Mandarin-speaking adults [3], yet relevant research on Mandarin-speaking children, which might probably shed light on the development of human vocal folds, remains limited.

Objectives: To probe into the characteristics of Mandarin-speaking children's voice by means of Electroglottography.

Methods: The participants were 106 Mandarin-speaking children aged 5 (53 boys and 53 girls). They were asked to read 3 sustained vowels [a], [i], and [u] and 1 nasal [m] in 5 different pitch levels ranging from low to high, each read twice with a duration of about 2 seconds. Real-Time EGG Analysis Model 5138 version 2.6.5 (based on MultiSpeech Model 3700 version 2.6.2) was employed to extract F0, OQ, and SQ. Data analysis was conducted using R version 4.4.2. Statistical analyses included Linear Mixed-Effects Model [4] [5].

Results: The results show the general pattern of the electroglottogram parameters in Mandarin native children's voice.

Above all, the average OQ of the children's voice is less than 50%, indicating that the open phase is shorter than the closed phase in their voice. Their SQ has an average of less than 180%.

Table 1. Descriptive statistics for the electroglottogram parameters in the children (N=106)

Parameters	M	SD
Pitch	316	48.2
Open quotient	47	10.9
Speed quotient	183	83.5

Furthermore, as for the relationships among the electroglottogram parameters, this study found that there was a negative correlation between OQ and F0 ($p<0.001$, Bonferroni corrected) as well as a negative correlation between SQ and F0 ($p<0.001$, Bonferroni corrected) in Mandarin children's voice.

Table 2. Model output for the relationships between pitch and the electroglottogram parameters

Predictors	Estimate	SE	df	t	p
(Intercept)	315.8	3.4	9.3	93.31	<0.001
Speed quotient	-5.6	0.1	630966.4	-81.171	<0.001
Open quotient	-1.692	0.1	630954	-25.495	<0.001
Speed quotient * Open quotient	-3.092	0.1	630951.5	-50.395	<0.001

Conclusions: This study found that the open phase is shorter than the closed phase in their voice, and that a higher F0 would result in both smaller OQ and smaller SQ in Mandarin native children's voice. These findings might assist in revealing the intrinsic relationships among various acoustic parameters in children's voice and the development of human vocal folds, which, hopefully, could be a reference for future research.

Acknowledgments: This research is funded by the Social Science Foundation of Beijing, China. Grant No: 24YYC015.

References:

- [1] Yokonishi, H., Imagawa, H., Sakakibara, K. I., Yamauchi, A., Nito, T., Yamasoba, T., & Tayama, N. (2016). Relationship of various open quotients with acoustic property, phonation types, fundamental frequency, and intensity. *Journal of Voice*, 30(2), 145-157.
- [2] Fang, T., Fant, G., Gao, F., & Gauffin, J. (1994). *Speech Science and Speech Technology*. Beijing: The Commercial Press.
- [3] Kong, J. (2001). *On language phonation*. China Minzu University Press.
- [4] lme4 : Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1-48.
- [5] lmerTest : Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H.B. (2017). lmerTest: Tests in linear mixed effects models. *R Journal*, 9(2), 2-32.

A Tool for Efficient HSV Rating based on the Voice-Vibratory Assessment with Laryngeal Imaging (VALI) form

Jonas Donhauser¹, Anne Schützenberger¹, Rita Patel², Michael Döllinger¹

¹ Division of Phoniatrics and Pediatric Audiology, Department of Otorhinolaryngology, Head and Neck Surgery, University Hospital Erlangen, Friedrich-Alexander University Erlangen-Nürnberg, Erlangen, Germany

² Department of Speech, Language, and Hearing Sciences, Indiana University, Bloomington, Indiana, USA

Keywords: HSV, rating, VALI

Abstract

Introduction & rationale:

In a clinical routine, endoscopy is used for inspection of the vocal folds to assess the healthiness of vocal fold motion and detect organic anomalies. For researchers, the analysis of High-Speed Video endoscopy (HSV) is of particular interest, as aperiodic motion is better resolved at higher frame rates. Besides HSV recordings, meaningful labels are required as ground-truth for various machine learning tasks, such as severity estimation.

A reliable visual-perceptual rating, which standardizes the rating of endoscopic recordings, is the Voice-Vibratory Assessment with Laryngeal Imaging (VALI) form (Poburka BJ, 2017). In a machine learning context, VALI features can either be used as individual labels or for combined HSV severity judgement.

Objectives:

Filling out the VALI form on paper is time consuming. Hence, to ease the HSV rating workflow, a software in the style of the VALI form was developed. By leveraging digital benefits, such as visual guidelines for scalar value specification, we aim to simplify and standardize HSV data set creation.

Explicit declaration of organic anomalies, as well as meta-features like overall severity and technical video quality, were added to the form to enhance the rating's usability for machine learning. Investigations on the application's inter-judge and intra-judge reliability will assess the reliability of digitized and newly incorporated features.

Furthermore, the relation of basic features to an overall HSV severity label will be investigated.

Methods:

A VALI like form was redesigned as digital application with faster HSV rating in mind. *PyQt* was chosen as framework to allow further individualization of the application by researchers who are familiar with the programming language *Python*. For each feature in the form an interactive widget was developed, enabling automatic computation of scalar values by clicking.

To assess the reliability of each implemented feature, a total of 200 HSV recordings, including 150 organic dysphonia recordings (50 polyps, 50 paresis, 50 atrophy), and 50 healthy patient recordings, will be labeled and evaluated statistically.

Results:

Results on the statistical inter-judge and intra-judge reliability are not available yet.

Conclusions:

An interactive application was developed as digitally enhanced software based on the VALI form. Through the digitalization, rating HSV recordings becomes faster, and thus eases HSV data set creation for machine learning in vocal fold research.

References:

- Poburka BJ, Patel R. (2017). Voice-Vibratory Assessment With Laryngeal Imaging (VALI) Form: Reliability of Rating Stroboscopy and High-speed Videoendoscopy. *J Voice*. doi:10.1016/j.jvoice.2016.12.003

Temporal Changes in Relative Fundamental Frequency in Continuous Speech: Contrasting Effects of State Vocal Fatigue and Vocal Fatigue Index Scores

Mark Berardi¹, Yixiang Gao², Gui DeSouza², Maria Dietrich^{3,4}

¹ Department of Communication Sciences and Disorders, University of Iowa, Iowa City, IA, USA

² Department of Electrical Engineering and Computer Science, University of Missouri, City, Columbia, MO, USA

³ Department of Speech, Language and Hearing Sciences, University of Missouri, Columbia, MO, USA

⁴ Department of Psychiatry and Psychotherapy, University Hospital Bonn, Bonn, NRW, Germany

Keywords: vocal fatigue, cluster analysis

Abstract

Introduction & rationale: Relative fundamental frequency (RFF) has emerged as a valuable acoustic measure for estimating laryngeal tension during speech production [1]. RFF measures the semitone change of glottal pulses during transitions from a vowel to a voiceless consonant and back to a vowel, providing insights into perceived vocal effort related to laryngeal tension. While RFF has been validated in studying various voice disorders and conditions [2] its relationship with vocal fatigue remains understudied. Recent advances in automated RFF analysis now enable the examination of larger data sets from continuous speech [3], allowing for more ecologically valid investigations of voice production patterns.

Objectives: This study aimed to determine the effects of both state vocal fatigue (acute fatigue developed during extended speech production) and self-reported vocal fatigue assessment (vocal fatigue as measured by the Vocal Fatigue Index) on RFF production in continuous speech. By examining these relationships, we sought to better understand how different measures of vocal fatigue assessment might influence laryngeal tension patterns during speech.

Methods: Eighty-two female participants (age range: 21-39 years, M = 24.3, SD = 3.3) were recruited, including both individuals with and without self-reported vocal fatigue but without phonotraumatic changes of their vocal folds as confirmed by laryngeal videostroboscopy. Self-report vocal fatigue was quantified using the first subsection of the Vocal Fatigue Index (VFI-1 [4], range: 0-28). For more information about inclusion and exclusion criteria and procedures see [5].

Participants were recorded reading two sentences containing target vowel-consonant-vowel (VCV) utterances 55 times each. The sentences were "The dew shimmered over my shiny blue shell again" and "Only we feel you do fail in new fallen dew," with target fricatives /ʃ/ and /f/ respectively. Recordings were made using a head-set microphone (AKG C520) in a sound isolation booth at 44.1 kHz with 16-bit resolution.

For analysis, the 55 trials were divided into three time periods: Early (trials 1-15), Mid (trials 21-35), and Late (trials 40-55). RFF instances were automatically extracted and analyzed using a previously validated automated algorithm [3]. Then an analysis pipeline employed machine learning techniques for feature extraction and clustering:

1. Feature Extraction:

- 20 RFF instance features were initially extracted (10 offset, 10 onset) for each VCV utterance
- Principal Component Analysis (PCA) was applied for dimensionality reduction
- The first three onset features explained the most variance, leading to removal of all offset features

2. Clustering Analysis:

- K-means clustering was implemented with silhouette score optimization
- Optimal number of clusters was determined by testing configurations from 2 to 10 clusters
- Each participant's samples were classified into clusters for each time period
- Cluster centers revealed distinct high and low RFF patterns

The proportion of samples in the low RFF cluster was computed for each participant across all time periods. Statistical analysis included normality testing (Shapiro-Wilk), homogeneity of variance testing (Levene's test), and mixed-design ANOVA with proper sphericity corrections.

Results: Table 1 summarizes the cluster centers for RFF Onset 1. There was about 1.1 ST difference between the clusters. This difference is similar to ones found in clinically relevant applications [2].

Statistical analysis revealed a significant main effect of Time ($F(2,188) = 11.25, p < 0.001$), indicating that RFF values changed significantly across the speaking task. Post-hoc analyses showed significant differences between Early vs. Late utterances (mean difference = 0.138) and Early vs. Mid utterances (mean difference = 0.117), but no significant difference between Mid and Late utterances.

The PCA revealed that onset features were more informative than offset features for characterizing RFF patterns, with the first three onset features explaining the majority of variance in the data. This finding suggests that voice onset transitions may be more sensitive indicators of vocal function than offset transitions.

Surprisingly, the main effect of Vocal Fatigue (VFI) was not significant ($F(1,94) = 1.23, p = 0.27$), suggesting that self-report vocal fatigue levels did not significantly interact with RFF generally. The interaction between Time and VFI was also non-significant ($F(2,188) = 0.097, p = 0.91$), indicating that the pattern of RFF changes over time was similar regardless of participants' baseline VFI-1 scores. Effect sizes were consistently small (all Cohen's $d < 0.2$), confirming minimal practical significance of self-report vocal fatigue on RFF measures. See Figure 1 for statistical summary.

Conclusions: This study provides evidence that onset RFF values significantly decrease during extended speech, particularly between early and later utterances, suggesting the development of state vocal fatigue. However, self-report vocal fatigue, as measured by the VFI, did not significantly influence RFF patterns. These findings suggest that RFF may be more sensitive to acute changes in vocal function during continuous speech than to perceived vocal fatigue. The current study demonstrates that 55 repetitions is effective to induce potential fatigue effects and fewer repetitions may be sufficient. This has important implications for the clinical use of RFF as a measure of vocal function and vocal fatigue, particularly in the context of continuous speech assessment.

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. *Journal of Speech, Language, and Hearing Research*, 53(5), 1220–1226. [https://doi.org/10.1044/1092-4388\(2010/09-0234\)](https://doi.org/10.1044/1092-4388(2010/09-0234))
- [2] McKenna, V. S., Vojtech, J. M., Previtera, M., Kendall, C. L., & Carraro, K. E. (2022). A scoping literature review of relative fundamental frequency (RFF) in individuals with and without voice disorders. *Applied Sciences*, 12(16), 8121. <https://doi.org/10.3390/app12168121>
- [3] Berardi, M.L., Tippit, E., DeSouza, G., & Dietrich, M. (2021, June 7-10). Automatic segmentation of relative fundamental frequency from continuous speech. [Paper presentation]. The 14th International Conference on Advances in Quantitative Laryngology, Voice and Speech Research (AQL), Bogotá, Colombia.
- [4] Nanjundeswaran, C., Jacobson, B. H., Gartner-Schmidt, J., & Verdolini Abbott, K. (2015). Vocal fatigue index (VFI): Development and validation. *Journal of Voice*, 29(4), 433–440. <https://doi.org/10.1016/j.jvoice.2014.09.012>
- [5] Gao, Y., Dietrich, M., & DeSouza, G. N. (2021). Classification of vocal fatigue using SEMG: Data Imbalance, normalization, and the role of Vocal Fatigue Index scores. *Applied Sciences*, 11(10), 4335. <https://doi.org/10.3390/app11104335>

Table 1. Cluster centers for RFF Onset 1 in semitones (RFF)

CLUSTER	EARLY UTTERANCES	MID UTTERANCES	LATE UTTERANCES
LOW ONSET RFF	1.65 ST	1.65 ST	1.63 ST
HIGH ONSET RFF	2.85 ST	2.78 ST	2.74 ST

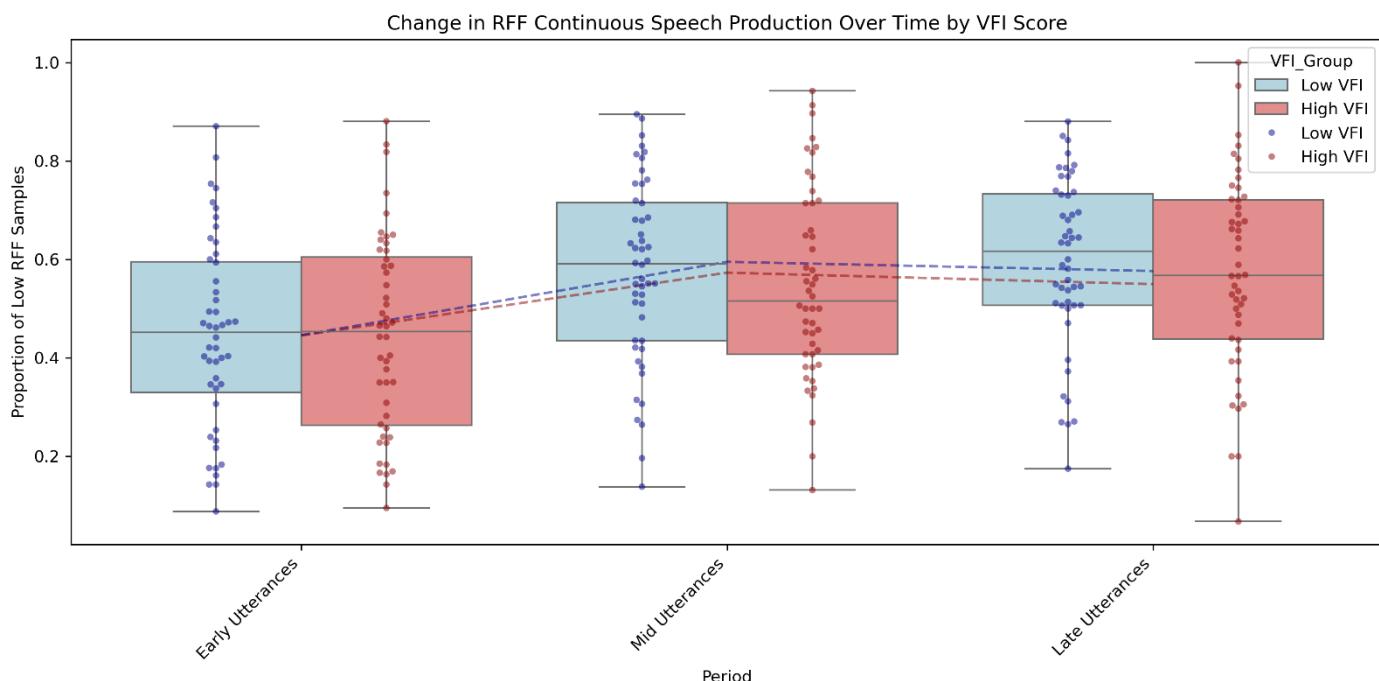


Figure 1. Relationship between high and low VFI-1 groups and proportion of low RFF samples across three time periods of continuous speech (Early: trials 1-15; Mid: trials 21-35; Late: trials 40-55).

Classification of Adductor Spasmodic Dysphonia with audio and glottal area waveform fusion

Hiroki Sato¹, David Crandall¹, Weslie Khoo¹, Rita Patel^{2,3}

¹ Luddy School of Informatics, Computing, and Engineering, Indiana University Bloomington, Bloomington, IN, USA

² Dept. Of Speech, Language and Hearing Sciences, Indiana University Bloomington, Bloomington, IN, USA

³ School of Medicine / Department of Otolaryngology Head and Neck Surgery, Indiana University Bloomington, Bloomington, IN, USA

Keywords: Deep Neural Networks; Glottal Area Waveforms; Vocal Pathology

Introduction:

An important goal for vocal pathology researchers is to develop methodologies for faster and more accurate pathological diagnosis in real-world settings. Assessing and correctly diagnosing spasmodic dysphonia is challenging due to its overlap with muscle tension dysphonia (MTD) and essential vocal tremor (Calà et al., 2023). Recent work (Merzougui et al., 2024; Fang et al., 2019; Hu et al., 2021) has examined automated classification of adductor spasmodic dysphonia (AdSD) and other vocal pathologies using artificial intelligence approaches, including deep learning models such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs) applied to acoustic features.

High-speed videoendoscopy (HSV) has allowed us to capture highly accurate data on vocal fold vibrations, including the glottal area waveform (GAW). Previous studies (Ahmad et al., 2012; Arias-Vergara et al., 2023) demonstrated that the GAW extracted from HSV, along with its Nyquist plot, is a robust and informative analysis tool for the vibratory patterns of vocal folds. Arias-Vergara et al. (2023) incorporated a shape-based parameter derived from the Nyquist plot as an input feature for a support vector machine classifier to distinguish functional dysphonia from healthy patients. In research on AdSD, Patel et al. (2011) visually analyzed vibratory motion waveforms from two-second HSV samples. Their findings identified oscillatory breaks and motion irregularities as key characteristics that differentiate AdSD from MTD, highlighting the usefulness of the vibratory features extracted from the GAW.

While previous work has shown the potential of GAW and machine learning in analytical studies on pathological voices, no work has explored integrating audio speech signals and GAWs for identifying AdSD using machine learning. In the machine learning communities, fusing multiple modalities, such as video and audio, has long been used for tasks such as action and activity recognition in video. More recently, Transformer models with attention mechanisms (Vaswani, et al., 2017) have become popular for combining evidence across multiple modalities and time scales. However, Transformers are known to require large amounts of training data and computation, and their potential application in voice research for AdSD remains largely unexplored.

Objectives:

This work builds on our current research on the detection and severity assessment of AdSD from audio by additionally integrating evidence from the GAW. We aim to investigate several deep learning techniques with different advantages and disadvantages to evaluate which works best in practice for AdSD identification from audio waveforms and GAW signals of sustained vowel phonation. We will explore two high-level approaches: (1) late-fusion, in which features are extracted from the audio and GAW signals separately before being combined for AdSD detection; and (2) early-fusion, in which both signals are modeled simultaneously to produce combined features, allowing interactions between the two signals to be considered. Specifically, we will investigate late- and early-fusion approaches with three specific types of neural network models: (1) CNNs, in which the audio and GAW are converted to two-dimensional spectrogram images, features are extracted from each spectrogram, and then the evidence from the two are combined; (2) RNNs, in which the audio and GAW signals are viewed as a sequence of observations, features are extracted from each sequence, and then the evidence is combined; and (3) Transformers, in which the signals are converted to spectrogram images and then a cross-attention mechanism is used to extract features from both images simultaneously to make a decision. These three models offer different trade-offs: Transformers are more powerful models that can naturally implement early fusion and make fewer assumptions about their inputs, but CNNs and RNNs are typically easier to train, requiring much less training data and computation. To better understand these trade-offs for AdSD detection, we will compare performance on validation data with metrics such as accuracy, precision/recall, and sensitivity/specificity.

Methods:

1. Data

Participants were recruited to the IU Vocal Physiology and Imaging Lab and IU Health Voice Center for an IRB-approved study. AdSD patients were included if they had not received Botox treatment in the past 12 months. Simultaneous grayscale HSV was conducted with an audio recording of sustained /i:/ phonation. HSV was captured at 4000 fps with a spatial resolution of 512 x 256 pixels using the Pentax Medical Model 9710 (Montvale, New Jersey). Acoustic recording occurred simultaneously using an omnidirectional microphone positioned 5 cm from the mouth at a sampling rate of 44.1kHz. The corresponding GAWs will be calculated using the Glottis Analysis Tool version 2020 (Lohscheller et al., 2007; Kist et al., 2021) (Erlangen, Germany). Data from 30 patients with AdSD and 30 healthy speakers will be analyzed.

2. Model

Using Mel filter bank spectrograms, we will extract time-frequency transformations from each audio signal and GAW. We will implement and evaluate at least one specific model architecture for each of the three model types mentioned above (CNNs, RNNs, and Transformers). For CNNs, we will use ResNet18 (He et al., 2016) with late-fusion: the model will consist of separate ResNet18 models to extract features from the audio and from the GAW, followed by a fully-connected layer to perform the classification. We will train the full model end-to-end and investigate using both pretrained and randomly-initialized weights. For the RNNs, we will use a similar two-stream late-fusion architecture but with Long-Short Term Memory (LSTM) (Sepp & Jürgen, 1997) models as the feature encoders. For the Transformers, we will investigate both late- and early-fusion approaches. For late-fusion, we will extract features using Audio Spectrogram Transformers (AST) (Gong et al., 2021), which support variable length input and can find correlations across different times and frequencies through the attention mechanism and then concatenate the audio and GAW features after the encoder. For early-fusion, we will use a Vision Transformer (ViT) (Dosovitskiy et al., 2021) with a cross-attention mechanism, which may allow it to extract features that incorporate evidence from both audio and GAW simultaneously.

For all experiments, we will use K-fold cross-validation to split the collected data into training and validation sets disjoint at the speaker level. This will prevent the model from overfitting to particular speakers' voices. Given the inherent stochasticity of deep learning, we will repeat all experiments 10 times and report means and standard deviations. We will implement our models in the Python programming language using Pytorch (Paszke, et al., 2019).

Results:

The results of the training and testing are not available yet.

Conclusion:

This study investigates the detection of AdSD through multimodal fusion of audio and GAW. For integrating the two types of evidence, we study two main techniques, late-fusion and early-fusion, through experiments with three different types of deep learning models: CNNs, RNNs, and Transformers. Since each of these approaches has different advantages and disadvantages, our study will produce new insights into how best to combine multimodal signals and deep learning to automatically and accurately identify AdSD.

References:

- Ahmad, K., Yan, Y., & Bless, D. (2012). Vocal fold vibratory characteristics of healthy geriatric females - Analysis of high-speed digital images. *Journal of Voice*, 751-759.
- Arias-Vergara, T., Döllinger, M., Schraut, T., Mohd Khairuddin, K. A., & Schützenberger, A. (2023). Nyquist Plot Parametrization for Quantitative Analysis of Vibration of the Vocal Folds. *Journal of Voice*.
- Calà, F., Frassineti, L., Manfredi, C., Dejonckere, P., Messina, F., Barbieri, S., . . . Cantarella, G. (2023). Machine Learning Assessment of Spasmodic Dysphonia Based on Acoustical and Perceptual Parameters. *Bioengineering*.
- Chen, C.-F., Fan, Q., & Panda, R. (2021). CrossViT: Cross-Attention Multi-Scale Vision Transformer for Image Classification. *International Conference on Computer Vision (ICCV)*, (pp. 347-356).
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., . . . Houlsby, N. (2021). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *ICLR*.
- Fang, S.-H., Tsao, Y., Hsiao, M.-J., Chen, J.-Y., Lai, Y.-H., Lin, F.-C., & Wang, C.-T. (2019). Detection of Pathological Voice Using Cepstrum Vectors: A Deep Learning Approach. *Journal of Voice*, 634-641.
- Gong, Y., Chung, Y.-A., & Glass, J. (2021). AST: Audio Spectrogram Transformer. *Interspeech*, (pp. 571-575).
- Hu, H.-C., Chang, S.-Y., Wang, C.-H., Li, K.-J., Cho, H.-Y., Chen, Y.-T., . . . Lee, O. K.-S. (2021). Deep Learning Application for Vocal Fold Disease Prediction Through Voice Recognition: Preliminary Development Study. *Journal of Medical Internet Research*.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Identity Mappings in Deep Residual Networks. *ECCV* (pp. 630-645).
- Kist, A. M., Gómez, P., Dubrovskiy, D., Schlegel, P., Kunduk, M., Echternach, M., . . . Döllinger, M. (2021). A Deep Learning Enhanced Novel Software Tool for Laryngeal Dynamics Analysis. *Journal of Speech, Language, and Hearing Research*, 1889-1903.
- Lohscheller, J., Toy, H., Rosanowski, F., Eysholdt, U., & Döllinger, M. (2007). Clinically evaluated procedure for the reconstruction of vocal fold vibrations from endoscopic digital high-speed videos. *Med Image Analysis*, 400-413.
- Merzougui, N., Korba, M. C., & Amara, F. (2024). Diagnosing Spasmodic Dysphonia with the Power of AI. *2024 ASU International Conference in Emerging Technologies for Sustainability and Intelligent Systems (ICETSIS)* (pp. 1042-1046). IEEE.
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., . . . Chintala, S. (2019). PyTorch: An imperative style, high-performance deep learning library. *Advances in Neural Information Processing Systems*.
- Patel, R. R., Liu, L., Galatsanos, N., & Bless, D. M. (2011). Differential vibratory characteristics of adductor spasmodic dysphonia and muscle tension dysphonia on high-speed digital imaging. *The Annals of Otology, Rhinology, and Laryngology*, 21-32.
- Sepp, H., & Jürgen, S. (1997). Long Short-Term Memory. *Neural Comput*, 1735-1780.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., . . . Polosukhin, I. (2017). Attention Is All You Need. *Neural Information Processing Systems* (pp. 6000-6010).

Speech Intelligibility in Speakers with Adductor Laryngeal Dystonia (AdLD)

Turley Duque¹, Jenny M. Vojtech¹, Katherine L. Marks², Laura E. Toles³, Cara Sauder⁴,
Tanya L. Eadie⁴, & Cara E. Stepp^{1,5,6}

¹ Department of Speech, Language, and Hearing Sciences, Boston University, Boston, Massachusetts, USA

² Department of Otolaryngology Head and Neck Surgery, Emory University School of Medicine, Atlanta, Georgia, USA

³ Otolaryngology – Head and Neck Surgery, University of Texas Southwestern Medical Center, Dallas, Texas, USA

⁴ Department of Speech and Hearing Sciences, University of Washington, Seattle, Washington, USA

⁵ Department of Otolaryngology – Head and Neck Surgery, Boston University School of Medicine, Boston, Massachusetts, USA

⁶ Department of Biomedical Engineering, Boston University, Boston, Massachusetts, USA

Keywords: speech perception, voice disorders, dysphonia

Abstract

Introduction & rationale: Adductor laryngeal dystonia (AdLD) is a rare neurological voice disorder. Individuals with AdLD have spasms in the muscles that adduct the vocal folds during phonation, resulting in a dysphonic voice that has intermittent voice stoppages and auditory-perceptual features of strain and roughness (Blitzer, 2010; Simonyan et al., 2021). This dysphonia negatively affects an individual's quality of life (Baylor et al., 2005), but more research needs to be done investigating if intelligibility (how well a speakers' message is understood) is also impacted in these individuals, especially in everyday noisy environments (Smith et al., 1998). It is also important to assess if there is a relationship between the severity of an individual's dysphonia and their intelligibility, as this could help guide treatment plans.

Objectives: This study investigated the difference in intelligibility of individuals with AdLD compared to those without AdLD (controls), and characterized the relationship between the overall severity of dysphonia of the AdLD speakers and their intelligibility.

Methods: Speakers were 44 individuals with AdLD and 44 age- and sex-matched controls. Each speaker read out a unique set of six Sentence Intelligibility Test sentences (Yorkston et al., 1996). Overall severity of dysphonia was assessed by five experienced speech-language pathologists specializing in voice using the Consensus Auditory Perceptual Evaluation-Voice (Kempster et al., 2009). For the intelligibility task, multi-speaker babble was added to each speech sample at a signal-to-noise ratio of one decibel to reduce ceiling effects (Abur et al., 2019). Five inexperienced listeners listened and orthographically transcribed all sentences, and intelligibility was calculated as the percentage of words matching the true transcription of the sentences. An analysis of variance was performed to determine the effects of group, stimulus (sentence length), and their interaction on intelligibility. A Pearson correlation coefficient was used to assess the strength of the linear relationship between intelligibility and overall severity of dysphonia in speakers with AdLD.

Results: Individuals with AdLD had significantly lower intelligibility ($M = 70.5$, $SD = 23.1$) than controls ($M = 84.9$, $SD = 9.8$) in noise. As sentence length increased, intelligibility significantly decreased ($p < .05$) for both the AdLD and control groups. Individuals with AdLD with more severe dysphonia had significantly lower intelligibility ($r = -0.35$, $p = .019$).

Conclusions: The current findings show that AdLD not only impacts voice quality, but also an individual's ability to be understood in noisy environments. The intelligibility of an individual with AdLD is lower than that of speakers without AdLD, and intelligibility decreases as their dysphonia increases in severity. The negative impacts of intelligibility should be considered when creating treatment strategies for individuals with AdLD.

References:

- Abur, D., Enos, N. M., & Stepp, C. E. (2019). Visual Analog Scale Ratings and Orthographic Transcription Measures of Sentence Intelligibility in Parkinson's Disease With Variable Listener Exposure. *American Journal of Speech-Language Pathology*, 28(3), 1222–1232. https://doi.org/10.1044/2019_AJSLP-18-0275
- Baylor, C. R., Yorkston, K. M., & Eadie, T. L. (2005). The consequences of spastic dysphonia on communication-related quality of life: A qualitative study of the insider's experiences. *J Commun Disord*, 38(5), 395–419. <https://doi.org/10.1016/j.jcomdis.2005.03.003>
- Blitzer, A. (2010). Spastic dysphonia and botulinum toxin: Experience from the largest treatment series. *European Journal of Neurology*, 17(s1), 28–30. <https://doi.org/10.1111/j.1468-1331.2010.03047.x>

- Kempster, G. B., Gerratt, B. R., Verdolini Abbott, K., Barkmeier-Kraemer, J., & Hillman, R. E. (2009). Consensus Auditory-Perceptual Evaluation of Voice: Development of a Standardized Clinical Protocol. *American Journal of Speech-Language Pathology, 18*(2), 124–132. [https://doi.org/10.1044/1058-0360\(2008/08-0017\)](https://doi.org/10.1044/1058-0360(2008/08-0017))
- Simonyan, K., Barkmeier-Kraemer, J., Blitzer, A., Hallett, M., Houde, J. F., Jacobson Kimberley, T., Ozelius, L. J., Pitman, M. J., Richardson, R. M., Sharma, N., Tanner, K., on behalf of the The NIH/NIDCD Workshop on Research Priorities in Spasmodic Dysphonia/Laryngeal Dystonia, Berke, G., Eadie, T., Greenlee, J., Hammer, M., Johns, M., Konczak, J., Ludlow, C., ... Stepp, C. (2021). Laryngeal Dystonia: Multidisciplinary Update on Terminology, Pathophysiology, and Research Priorities. *Neurology, 96*(21), 989–1001. <https://doi.org/10.1212/WNL.00000000000011922>
- Smith, E., Taylor, M., Mendoza, M., Barkmeier, J., Lemke, J., & Hoffman, H. (1998). Spasmodic dysphonia and vocal fold paralysis: Outcomes of voice problems on work-related functioning. *J Voice, 12*(2), 223–232. [https://doi.org/10.1016/s0892-1997\(98\)80042-8](https://doi.org/10.1016/s0892-1997(98)80042-8)
- Yorkston, K. M., Beukelman, D. R., & Tice, R. (1996). *Speech Intelligibility Test* [Computer software]. Tice Technologies.

The Effect of Intentional Voice Quality Alteration on Mental Effort Captured via Pupillometry and via the NASA-TLX and Borg CR-10 scales

Eva van Leer¹, Lucas Ribas², Madison Won¹, Tianqi Luan³

¹ Department of Communication Sciences & Disorders, Georgia State University, Atlanta, GA, USA ² Department of Applied Linguistics & ESL, Georgia State University, Atlanta, GA, USA

³ Department of Computer Science, Georgia State University, Atlanta, GA, USA

Keywords: voice therapy, pupillometry, mental effort

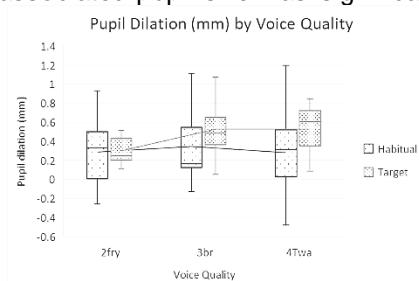
Abstract

Introduction & rationale: Mental effort- the amount of cognitive work done to complete a task (Mulder, 1986)- is known to increase when individuals intentionally alter their voice quality in artistic voice training (Lee et al., 1994; Verdolini, 1997) and voice therapy (van Leer & Connor, 2010). This increase in mental effort is a documented aspect of motor learning that reduces with extensive practice (Lee et al., 1994). Because cognitive resources are finite (Vinney & Turkstra, 2013), the mental demands of voice homework practice can present a barrier to practice (e.g. “sometimes I don’t have mental energy to focus on my voice”) and to the application of the vocal target (e.g. “I can’t focus on the conversation and my voice at the same time.”) (van Leer & Connor, 2010). While mental effort therefore plays an important role in vocal skill acquisition, the topic has received almost no empirical investigation. Many objective and subjective measurement approaches are available. Because the pupil dilates when mental effort is increased (Darwin, Charles, 1872; Mulder, 1986), pupillometry presents an objective measure. Alternatively, self-report scales can be used to measure mental effort subjectively.

Objectives: The purpose of this study was threefold: to examine the effect of intentional voice quality alteration on mental effort using both 1) objective cognitive pupillometry methods and 2) subjective self-report scales, and to 3) examine the association between these two measurement approaches.

Methods: Eleven participants without formal speaking-voice training were instructed to produce three voice qualities: breathy, fry and twang. For recording of pupil diameter during voice production they were fitted with iMotion SMI eye tracking glasses and turned toward a non-reflective grey wall in a windowless room with a constant luminance level. Three separate pupillometry recordings were made: one for fry, one for breathy, and one for twang. In each recording, participants alternated 30 second periods each of 1) silence 2) counting aloud in their habitual voice 3) silence, and 4) counting aloud in the target; this sequence was repeated three times and concluded with a 30 second silent period yielding a total of 6.5 minutes per recording. After each recording, participants completed the Borg CR-10 (Vinney et al., 2023) and NASA-Mental Demand subscale (Hart, 1988), rested for 15 minutes, and completed the next recording. Pupillometry data was pre-processed for blink removal and processed for baseline removal (i.e. removal of pupil size during silence). Paired t test were applied to examine the difference between pupil size in target voice versus habitual voice production, and the difference between self-reported mental effort in these conditions. Spearman’s rho was calculated to examine the association between self-report data and baseline-corrected pupil diameter.

Results: For all three intentionally altered voices- fry, breathy and twang- associated pupil size was significantly larger than for habitual voice production ($p < .001$), with very large effect sizes (Cohen’s d exceeding 1) as shown in Fig 1 below. Likewise, mental effort ratings on both the Borg CR-10 and NASA-TLX were significantly higher for the three target voice qualities than for habitual voice production ($p < .001$), with very large effect sizes exceeding 1. There was a significant moderate positive association between pupil dilation in millimeter and mental effort ratings on both the Borg CR-10 ($r = .56$) and NASA-TLX ($r = .54$). Exploratory investigation further suggested that twang and breathy voice were mentally more demanding than fry: physical effort and affect may play a role in explaining this preliminary finding.(Van Mersbergen & Payne, 2021) Findings are consistent with those showing a pupillometry and self-report difference between habitual speech production compared to a target speech production (i.e. clear speech).(Ishikawa, Coster, et al., 2023; Ishikawa, Li, et al., 2023)



Conclusions: As hypothesized, intentional voice quality change yielded both significant increases in pupil dilation and in participant-perceived mental effort, and these measures are associated with each other. Findings are fundamental to our understanding of the mental demands of therapeutic and artistic voice training.

References:

Darwin, Charles. (1872). *The expression of the emotions in man and animals*. John Murray.

- Hart, S. G. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Human Mental Workload/Elsevier*. <https://www.sciencedirect.com/science/article/pii/S0166411508623869>
- Ishikawa, K., Coster, E., Li, H., & Orbelo, D. (2023). *The use of pupillometry to study mental demand during speech production in noise*. <https://osf.io/preprints/psyarxiv/s9dph/>
- Ishikawa, K., Li, H., & Coster, E. (2023). The Effect of Noise on Initiation and Maintenance of Clear Speech and Associated Mental Demand. *Journal of Speech, Language, and Hearing Research*, 66(11), 4180–4190. https://doi.org/10.1044/2023_JSLHR-23-00157
- Lee, T. D., Swinnen, S. P., & Serrien, D. J. (1994). Cognitive Effort and Motor Learning. *Quest*, 46(3), 328–344. <https://doi.org/10.1080/00336297.1994.10484130>
- Mulder, G. (1986). The Concept and Measurement of Mental Effort. In G. R. J. Hockey, A. W. K. Gaillard, & M. G. H. Coles (Eds.), *Energetics and Human Information Processing* (pp. 175–198). Springer Netherlands. https://doi.org/10.1007/978-94-009-4448-0_12
- van Leer, E., & Connor, N. P. (2010). Patient perceptions of voice therapy adherence. *Journal of Voice*, 24(4), 458–469.
- Van Mersbergen, M., & Payne, A. E. (2021). Cognitive, emotional, and social influences on voice production elicited by three different Stroop tasks. *Folia Phoniatrica et Logopaedica*, 73(4), 326–334.
- Verdolini, K. (1997). Principles of skill acquisition applied to voice training. *Hampton, M., & Acker, B. The Vocal Vision: Views on Voice By*, 24, 65–80.
- Vinney, L. A., Tripp, R., Shelly, S., & Gillespie, A. (2023). Indexing Cognitive Resource Usage for Acquisition of Initial Voice Therapy Targets. *American Journal of Speech-Language Pathology*, 32(2), 717–732. https://doi.org/10.1044/2022_AJSLP-22-00197
- Vinney, L. A., & Turkstra, L. S. (2013). The role of self-regulation in voice therapy. *Journal of Voice*, 27(3), 390-e1.

A total variation based semi-automatic high-speed video denoising pipeline

Benjamin Peschel¹, Tony Schelhorn², Ulrich Hoppe², Rosa Uhl², Moritz Bingold¹, Michael Döllinger¹

¹Division of Phoniatrics and Pediatric Audiology, Department of Otorhinolaryngology, Head and Neck Surgery, University Hospital Erlangen, FAU Erlangen-Nürnberg, Germany

²Division of Audiology, Department of Otorhinolaryngology, Head and Neck Surgery, University Hospital Erlangen, FAU Erlangen-Nürnberg, Germany

Keywords: Image Processing, High-Speed Video, Endoscopy

Abstract

Introduction & rationale:

Stroboscopy is still the standard in vocal fold imaging for clinical assessment, even though the advantages of high-speed video endoscopy (HSV) are becoming more evident (Zacharias et al. 2018). Accessing the vocal folds via endoscope is either possible orally with a rigid endoscope or nasally via a flexible one. The main advantage of nasal endoscopy is less impediment to natural speech production (Yousef et al. 2022). From a technical perspective the drawbacks of using nasal endoscopy in conjunction with HSV are twofold. Firstly, the optical fibers inside the endoscope are structured in a hexagonal pattern, resulting in a characteristic noise pattern, see Figure 1 b). Secondly, only a fraction of the light from the light source reaches the end of the endoscope. Moreover, emitting too much light from the small head of the endoscope can cause burns. These limitations become amplified by the inherently small exposure time of HSV. As a result, the signal-to-noise ratio (SNR) can become dominated by the inherent camera noise (Wang et al. 2019) as seen in Figure 1 a). To maximize the amount of light hitting the camera sensor, the endoscope has to be as close to the vocal folds as possible. Hence, the SNR is highly dependent on the examinee's tolerance to having a foreign object inserted through the nose and the skill of the examiner with the endoscope. Increasing the video's SNR during post processing can therefore reduce the stress to the participant and the skill level required to capture HSV recordings of the vocal folds.

Objectives:

We aim to construct a HSV enhancement pipeline. Through this we will lower the barrier of entry on HSV research. We will construct a method with few tunable parameters, that are easily interpretable and mainly dependent on previously known features like the required frame rate, the camera and the endoscope used.

Methods:

On the one hand both HSV noise as well as organic movement show high frequency temporal components. On the other hand, organic movement exhibits a higher temporal correlation between frames when compared to the HSV specific noise, which is almost independent between frames. We therefore solved for each pixel over time the convex total variation (Rudin et al. 1992) problem using a primal dual minimization algorithm (Chambolle et al. 2011).

To remove the hexagonal pattern, we employed a circular spatial low pass filter with a cutoff frequency being half the density of fiber cables inside the endoscope. We estimated the distance between optical fibers automatically from peaks in the frequency domain of the frames (Winter et al. 2006). Finally, histogram equalization was used on each frame.

We tested the method on a data set consisting of 31 participants (mean age $53,6 \pm 23$ years, 16 male / 15 female) phonating the phoneme /a/. For each participant, 3 videos consisting of 50 ms phonation were randomly picked from 30 existing videos. Two flexible endoscopes were used, the ENF-GP2 (Olympus, Shinjuku, Japan) and the Rhino-Pharyngo-Laryngo-Fiberskop (Karl Storz, Tuttlingen, Germany). The endoscope was directly connected to a light source (LED 300, Karl Storz, Tuttlingen, Germany) and high-speed camera Phantom v2511 (Ametek, USA, Berwyn, PA) mounted on a movable camera arm.

The videos were captured with a spatial-temporal resolution of 384 x 288 pixels and 10 kHz respectively. The endoscopes were operated by two medical students. Neither had any experience in nasal endoscopy prior to a one-week internship in our division in preparation for the measurements. To judge the quality of the enhancement, we use the reference-free NIQE score model (Mittal et al. 2012), modeling the statistical distance to a reference data set. As a reference, 1024 randomly chosen images from the BAGLs (Gomez et al. 2020) data set were used. This data set offers a good quality reference of HSV data, mainly captured at or below 4 kHz and with an oral endoscope with sufficient light passing through. Nevertheless 92 images were manually removed, because of noise corruption. The two-sided Wilcoxon signed rank test was employed with the null hypothesis of no change in NIQE scores, a confidence level of $\alpha = 0.05$ was chosen. Because of the large sample size, we employed an asymptotic approximation using the python package SciPy (Virtanen et al. 2020).

Results:

Subjectively, the characteristic HSV noise could be removed, see Figure 1 b), without smoothing the glottis contour too much, Figure 1 c). Objectively, the average NIQE score improved from 34.1 (± 22.2) before processing to an average of 10.8 (± 1.5). The p-value of the asymptotic approximation of the Wilcoxon signed rank test was 0.0. The null hypothesis can therefore be rejected.

Conclusions:

The constructed method has been shown to significantly improve the SNR of HSV videos with nasal endoscopy, proxied by the NIQE score. This can potentially reduce stress on participants. Furthermore, it increases the number of usable videos made during an examination and reduces the skill required by the examiner. Moreover, the method has simple interpretable parameters. Total variation only has one parameter, representing the noise level. The remaining parameters, the window function and the kernel size of the filter, both have an extensive body of literature that can be consulted e.g., Tan et al. 2018. Nevertheless, the Gaussian model underlying the NIQE scores is limited. NIQE should therefore only be seen as proof of concept. Additional test metrics to validate the model are needed. The next step in our approach will be to calculate the fundamental frequency from the raw and the processed video and match it with simultaneously recorded audio data. We will also explore at what frame rate the model breaks down. Last but not least, further tests with different hardware are required.



Figure 1 a: Raw frame, 10 kHz, increased brightness

Figure 1 b: After temporal total variation denoising

Figure 1 c: After temporal total variation denoising and spatial filtering

References:

- Chambolle, A., & Pock, T. (2011). A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of mathematical imaging and vision*, 40, 120-145.
- 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), 186.
- Mittal, A., Soundararajan, R., & Bovik, A. C. (2012). Making a “completely blind” image quality analyzer. *IEEE Signal processing letters*, 20(3), 209-212.
- Rudin, L. I., Osher, S., & Fatemi, E. (1992). Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena*, 60(1-4), 259-268.
- Tan, L., & Jiang, J. (2018). *Digital signal processing: fundamentals and applications*. Academic press.
- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., ... & Van Mulbregt, P. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature methods*, 17(3), 261-272.
- Wang, W., Chen, X., Yang, C., Li, X., Hu, X., & Yue, T. (2019). Enhancing low light videos by exploring high sensitivity camera noise. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 4111-4119).
- Winter, Christian, et al. "Automatic adaptive enhancement for images obtained with fiberoptic endoscopes." *IEEE Transactions on Biomedical Engineering* 53.10 (2006): 2035-2046.
- Yousef, A. M., Deliyski, D. D., Zacharias, S. R., de Alarcon, A., Orlikoff, R. F., & Naghibolhosseini, M. (2022). A deep learning approach for quantifying vocal fold dynamics during connected speech using laryngeal high-speed videoendoscopy. *Journal of Speech, Language, and Hearing Research*, 65(6), 2098-2113.
- Zacharias, S. R., Deliyski, D. D., & Gerlach, T. T. (2018). Utility of laryngeal high-speed videoendoscopy in clinical voice assessment. *Journal of Voice*, 32(2), 216-220.

PODIUM SESSION 8

Arytenoid Adduction and Medialization: Resolving Height Asymmetry and Posterior Gap to Optimize Vocal Efficiency

Liran Oren^{1,2}, Charles Farbos De Luzan^{1,2}, Jacob Michaud-Dorko², Ephraim Gutmark^{3,1}, Greg Dion^{1,2}

¹ Department of Otolaryngology, University of Cincinnati, Cincinnati, Ohio, USA

² Department of Biomedical Engineering, University of Cincinnati, Cincinnati, Ohio, USA

³ Department of Aerospace Engineering, University of Cincinnati, Cincinnati, Ohio, USA

Keywords: Flow-Structure-Interaction, Vocal Efficiency, Unilateral Vocal Fold Paralysis,

Abstract

Introduction & rationale: Unilateral vocal fold paralysis (UVFP) is a condition in which one vocal fold cannot move toward the other. Often secondary to recurrent laryngeal nerve injury, it results in significant muscle atrophy and incomplete glottal closure (glottal insufficiency). This impairment leads to a breathy, weak voice, vocal fatigue, and reduced intelligibility, significantly affecting a person's quality of life. The primary surgical treatment for UVFP is Thyroplasty Type I (TT1), which focuses on repositioning the paralyzed vocal fold and adding bulk to the paraglottic space to improve glottic closure. However, this procedure often addresses the anterior and mid-membranous portions of the vocal fold more effectively than the posterior region, sometimes leaving a persistent posterior glottic gap (Woo et al., 2001). This residual gap can impair vocal efficiency (VE) by allowing airflow leakage and reducing the conversion of aerodynamic energy into acoustic energy. Additionally, asymmetry in the height of the vocal processes following medialization can further disrupt VE. These effects on VE contribute to inconsistent voice outcomes, leading many laryngologists to recommend incorporating an arytenoid repositioning procedure, such as arytenoid adduction (AA), to optimize surgical results (Courrey, 2012).

The interaction between airflow and tissue, known as flow-structure interaction, shapes intraglottal flow behavior and affects the forces acting on the glottal walls. Our previous work identified intraglottal flow separation vortices (FSV) as an important driver of VE (Cohen et al., 2024; Farbos de Luzan et al., 2020; Maddox et al., 2023). These vortices generate negative pressures that increase vocal fold closing speed and efficiency (Jiang et al., 2023). Further studies demonstrated that infraglottal medialization more effectively replicates the natural stiffness gradient, leading to superior VE compared to glottal medialization. Using excised canine larynx models, we showed that medializing the infraglottal region improves VE more effectively than medializing the fold itself (Oren et al., 2024).

The current study aims to clarify how AA improves VE in cases where asymmetry in the height of the vocal processes occurs after medialization, with and without a posterior gap; i.e., whether the enhanced VE with AA is due to closing the posterior gap, correcting the height asymmetry, or the combined effect of both. As these conditions frequently occur in patients, understanding their individual and combined contributions will help resolve the clinical uncertainty surrounding the indications for arytenoid adduction.

Objectives: This study aims to evaluate how AA impacts VE, FSV, and acoustic measures by addressing vocal process height asymmetry and posterior gap presence in cases of UVFP following medialization.

Methods: The study was conducted using four excised canine larynges. UVFP was simulated by medially adducting the "non-paralyzed" fold through translation of its vocal process, while the paralyzed fold remained in its natural resting position. A laryngologist performed all surgical manipulations using Silastic implants. The first case established the baseline, where the implant was carefully placed through a thyroplasty window in the thyroid cartilage on the paralyzed fold to achieve symmetrical adduction without a posterior gap (Fig 1a). We then considered four cases by inserting prongs into the vocal process of the paralyzed fold. The first two cases involved displacing the fold 1 mm and 2 mm inferiorly from the baseline position while maintaining no posterior gap. The next two cases introduced lateral displacements of 1 mm and 2 mm, creating both a posterior gap and a 2 mm height asymmetry (Fig 1b). After completing these asymmetrical cases, a suture was used to tighten and rotate the arytenoid. This final step simulated the effect of adding AA to a TT1 procedure. All of these six cases were simulated in each larynx.

Phonation for each case was induced at low and high levels of subglottal pressure by controlling the conditioned airflow supplied to the larynx. Acoustic measurements were recorded using a microphone positioned approximately 30 cm lateral and superior to the glottal exit. VE was calculated using its classic definition as the ratio of acoustic power to

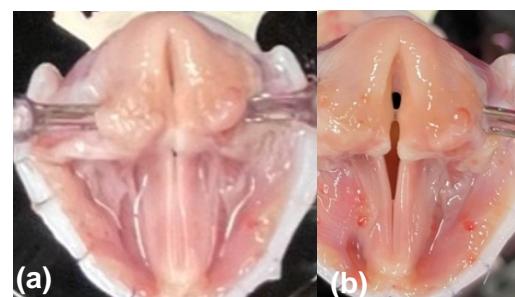


Figure 1. Simulated images of adducted vocal folds. a) Baseline case - symmetrical adduction without a posterior gap. b) Most detrimental case – 2 mm height asymmetry with 2 mm posterior gap.

aerodynamic power (Schutte, 1980). Intraglottal flow velocity field measurements were taken using particle image velocimetry.

Results: The preliminary results indicate that VE is significantly impacted when a TT1 procedure creates both height asymmetry and a posterior gap, and that adding AA can enhance VE (Fig. 2). The highest VE was observed when the vocal folds were symmetrically adducted without a posterior gap (i.e., baseline). VE progressively declined as height asymmetry and posterior gap size increased, with their combined presence following TT1 being the most detrimental. Introducing AA in this condition led to improved VE at both levels of subglottal pressure.

The effect of AA was only assessed in the most detrimental case. While AA effectively closed the posterior gap and mitigated glottal asymmetry, we observe that VE remained lower than in the baseline case. This comparison should be interpreted with caution, as the study was designed to simulate intraoperative conditions where TT1 leads to both asymmetry and a persistent gap. The resulting reduction in VE contributes to increased vocal fatigue, a common complaint among UVFP patients (Maddox et al., 2023), and suggest that adding AA could enhance VE by addressing both asymmetry and the posterior gap.

The relationship between height asymmetry, posterior gap, and FSV offers insight into the observed VE changes. FSV strength, quantified by circulation magnitude, was highest in the baseline case and decreased proportionally with increasing height asymmetry and gap size. The addition of AA increased circulation strength, but compared to baseline, the recirculation zones were smaller, resulting in lower overall circulation values. The reduction in circulation strength suggests a weaker negative pressure force acting on the glottal walls, which may slow vocal fold closure and impact VE.

Conclusions: The systematic variations demonstrate a progressive decline in VE as height asymmetry and posterior gap size increase. Additionally, the findings confirm that AA improves VE by correcting both height asymmetry and closing the posterior gap, with its effect being proportional to the severity of these initial deficits. By providing a continuous and measurable understanding of how VE is influenced by laryngeal asymmetries and medialization strategies, this study offers stronger predictive insights for optimizing surgical interventions.

References:

- Cohen, O., de Luzan, C. F., Michaud-Dorko, J., Howell, R. J., Dion, G. R., & Oren, L. (2024). Infraglottal Medialization: Increasing Vocal Fold Stiffness Gradient for Improved Vocal Efficiency. *Journal of Voice*.
- Courey, M. (2012). Management of Unilateral Vocal Fold Paralysis/Paresis: A Laryngologist's Perspective. *Perspectives on Voice and Voice Disorders*, 22(3), 121-130.
- Farbos de Luzan, C., Maddox, A., Oren, L., Gutmark, E., Howell, R. J., & Khosla, S. M. (2020). Impact of Vertical Stiffness Gradient on the Maximum Divergence Angle. *The Laryngoscope*.
- Jiang, W., Zheng, X., Farbos de Luzan, C., Oren, L., Gutmark, E., & Xue, Q. (2023). The Effects of Negative Pressure Induced by Flow Separation Vortices on Vocal Fold Dynamics during Voice Production. *Bioengineering*, 10(10), 1215.
- Maddox, A., Oren, L., Farbos de Luzan, C., Howell, R., Gutmark, E., & Khosla, S. (2023). An Ex-vivo Model Examining Acoustics and Aerodynamic Effects Following Medialization With and Without Arytenoid Adduction. *The Laryngoscope*, 133(3), 621-627.
- Oren, L., Maddox, A., Farbos de Luzan, C., Xie, C., Howell, R., Dion, G., Gutmark, E., & Khosla, S. (2024). Acoustics and aerodynamic effects following glottal and infraglottal medialization in an excised larynx model. *European Archives of Oto-Rhino-Laryngology*, 281(5), 2523-2529.
- Schutte, H. K. (1980). *The efficiency of voice production*. Kemper Groningen:.
- Woo, P., Pearl, A. W., Hsiung, M. W., & Som, P. (2001). Failed medialization laryngoplasty: management by revision surgery. *Otolaryngology—Head and Neck Surgery*, 124(6), 615-621.

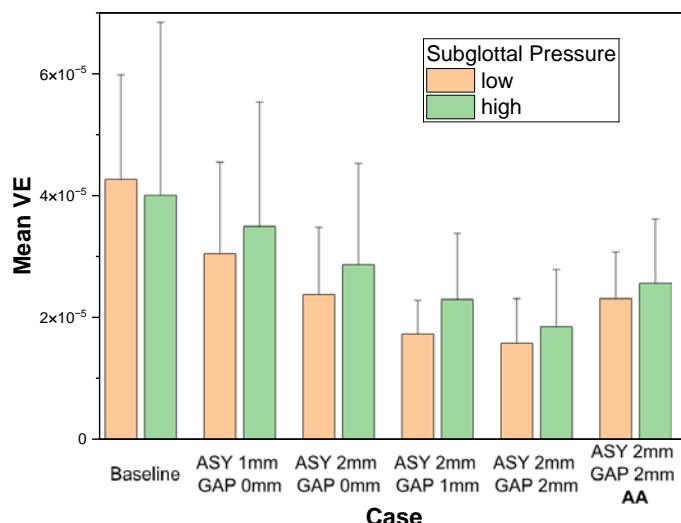


Figure 2. Mean vocal efficiency (VE) calculated from the four larynges as a function of the different adduction conditions.

Our results align with previous findings (Maddox et al., 2023), and suggest that adding AA could enhance VE by addressing both asymmetry and the posterior gap.

A Microengineered Vocal Fold Model for High-Throughput Screening Applications

Leila Donyaparastlivari¹, Scott Thomson², Chen Shen³, Amir K. Miri¹

¹ Department of Mechanical Engineering, New Jersey Institute of Technology (NJIT), Newark, NJ, US

² Department of Mechanical Engineering, Brigham Young University, Idaho, Idaho, ID, US

³ Department of Mechanical Engineering, Rowan University, Glassboro, NJ, US

Keywords: high-throughput screening, multiscale vibration, hybrid modeling.

Abstract

The recent advancements in microfabrication technologies have enabled the creation of intricate and highly controlled mimetic models with potential screening applications and an understanding of disease mechanisms. In the case of vocal fold tissue, such a model allows for replicating physiological loading onto cells. In this work, we aimed to scale down the geometry of the vibrating EPI model to induce sub-millimeter oscillations (as a μ -EPI model). Three different sizes of *in vitro* testbeds were specified for three vocal fold replicas: (i) 100%, (ii) 50%, and (iii) 25% of the original size. We used an established silicone elastomer, Ecoflex, and gelatin-based hydrogel to study the vibration pattern onto a synthetic EPI model. First, we optimized the setup based on phonation pressure. Then, we included one space for gelatin encapsulation. We measured the mechanics of vocal folds during finite strains under tension, compression, and shear loadings to study the impact of gelatin. The geometrical and mechanical parameters measured during the experiments emphasized the effects of the vocal fold geometry and composition on the vibration response and pressure threshold. Our results will be used to place a cell-laden hydrogel model into the miniaturized model.

Introduction & Rationale:

Voice disorders significantly impact the quality of life for professional users and singers. The voice production in the larynx results from flow-induced vibrations of the vocal folds, involving a complex interplay of flow dynamics, tissue oscillations, and acoustics [1]. Understanding the disease mechanisms and predicting the tissue-drug interactions require proper tools. Studying phonation-related disease mechanics in animal models has been challenging due to the differences between human and animal phonations. A mimetic vocal fold model is needed to decouple non-phonatory and phonatory factors under controlled conditions. Synthetic, self-oscillating vocal fold models have been used to replicate the native vibration patterns of human vocal folds, enabling in-depth studies of voice mechanics [2]. Scaling down the model and including cell-related compartments make them suitable for high-throughput applications. When used in physiological-relevant geometries, silicone-based polymers (Ecoflex) [3] can create self-sustained vocal fold vibrations. Two well-known models are the EPI [4] and MRI-based models [5]. In this work, we designed and assembled a bioreactor following the EPI model, which is made of Ecoflex and filled with a hydrogel-based mass.

Methods:

The design is shown in **Fig. 1**, and it includes a body with a tubing system, a Plexiglass base, and vocal fold pairs. The EPI model was scaled by a 1:4 ratio (μ -EPI) to match the dimensions of their vocal folds (**Fig. 1A-C**). A compressor (Senor air compressor, PC 1010), a pressure transducer (100PSI Pressure Transducer), a pressure gauge (IP54 Stainless Steel), and an airflow meter (MF5700 series airflow meter) were used to make the setup in **Fig. 1E**. A high-resolution pressure transducer (Omega) was mounted ~ 80 mm below the folds to measure the supraglottal pressure. The transducer was connected to a data acquisition system (MSIP-REM-NATUSB6001) for data collection. Dynamic supraglottal pressure was also monitored using a digital microphone (SLM-25 professional sound level meter)

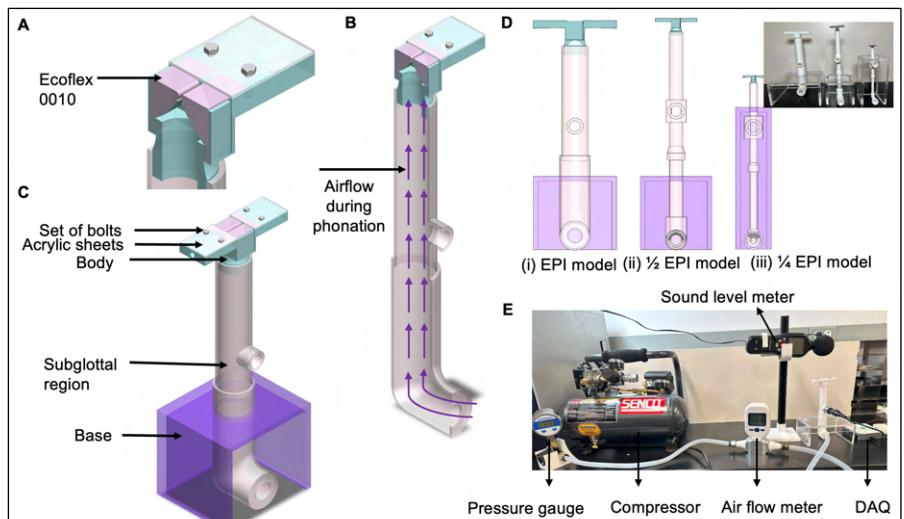


Figure 1. A) The body of the bioreactor with Ecoflex replicas inside the body, (B) A cut view for the bioreactor to show the airflow pathway, (C) the whole bioreactor with a base for holding the air-flow system and perfusion encasement, (D) the three main sizes of (i) EPI, (ii) 50%EPI, and (iii) 25%EPI with scalable chambers, and (E) the complete bioreactor setup made for this project (Miri Lab).

~ 80 mm above the folds. Data from the microphone was processed using the software (NoiseLogger Communication

Tool). The data for the pressure measurements of the synthetic model (via an Ecoflex:thinner ratio) are shown in **Fig. 2A-C**. The same procedure was applied for all EPI sizes. We used the Reynolds number as our criteria for scaling down. Instron Universal Tester and TA Instruments Rheometer were used to measure the physical properties of the structure (stiffness). We then measured the pressure threshold in the subglottal region connected to an airflow system at its base, which was monitored through a DAQ system to optimize the phonatory

response of the synthetic models. We put a threshold of 100 Hz vibration as a possible output. The results are shown in **Fig. 2D-E**. In the final step, the hydrogel samples were positioned at the cylinder's upper section to facilitate direct contact with airflow from the compressor. A high-speed camera (Fastcam Mini UX100) was placed above the samples to capture their movement during phonation, enabling relating the vibration frequencies and pressure values (*not shown here*). We encapsulated a GelMA capsule (i.e., 100-200 μ L in volume) into the model, making perfusions and observing the phonatory behavior of the model via a high-speed camera. The hydrogel was made of GelMA and a human vocal fold fibroblast before being placed within the plastic model (*there is a pending Patent on this work, submitted at NJIT*).

Results:

The physical properties of our material system showed a range of elastic moduli from 1 kPa to 1 MPa, as the literature requires an elasticity of around 50 kPa and a viscous modulus of around 20 kPa to mimic the native vocal fold tissue. We have optimized the Ecoflex:thinner ratio to reach the desired modulus and viscosity for the phonatory behavior (i.e., frequency and amplitude). There is a linear relation between airflow rate and subglottal pressure for all three cases, while smaller size led to a higher pressure with a ratio of ~ 40:1 for 1:4 size reduction. The linear response of the pressure vs flow rate is shorter in higher pressure values for a smaller model. Our results show the two folds' stable, self-sustaining vibration (making contact forces) under 2 kPa air pressure (see **Fig. 3**). We then tested the encapsulation of acellular hydrogel into the model for different shapes and locations. In our preliminary tests, we placed the GelMA models into the synthetic replicate via (i) molding and (ii) injection. It was found that the molding method gave more repeatable models (i.e., a lower standard deviation) and impaired surface properties (i.e., no clear vibration pattern). We used a syringe injection (20G) to place pre-warmed GelMA solutions into the mold. The phonatory function shows the vibration frequency's dependency on the gel's volume and the bulk stiffness. We then evaluated the cell viability and functionality of the hydrogel model under vibrations (*not shown here*).

Conclusions:

We designed and developed the prototype of a miniaturized vocal fold phonatory model with the capacity to encapsulate biological cells under physiological loading. We showed the capacity of our modeling approach to simulate self-sustained oscillations and hydrogel encapsulation capacity. The next step of this project is to optimize the cell capacity and gel position within the replicate to have a long-term perfusion of culture media. The stability of the model towards the 14-day time point is a key criterion for the screening application. The final step is to add polydimethylsiloxane (PDMS) to our system for scale-up and commercialization.

References:

1. M€uller, J. P. (1839). *Uber Die Compensation Der Physischen Kr € afte Am € Menschlichen Stimmorgan: Mit Bemerkungen Uber Die Stimme Der € Saugetiere, V € ogel Und Amphibien* (On the Compensation of the Physical € Forces in the Human Vocal Organ: With Remarks on the Voice of Mammals, Birds and Amphibians) (A. Hirschwald, Berlin).
2. Probst, Judith, et al. "Acoustic and aerodynamic coupling during phonation in MRI-based vocal tract replicas." *Applied Sciences* 9.17 (2019): 3562.
3. Latifi, Neda, et al. "A flow perfusion bioreactor system for vocal fold tissue engineering applications." *Tissue Engineering Part C: Methods* 22.9 (2016): 823-838.
4. Schoder, Stefan, et al. "A benchmark case for aeroacoustic simulations involving fluid-structure-acoustic interaction transferred from the process of human phonation." *Acta Acustica* 8 (2024): 13.
5. Wu, Liang, and Zhaoyan Zhang. "A parametric vocal fold model based on magnetic resonance imaging." *The Journal of the Acoustical Society of America* 140.2 (2016): EL159-EL165.

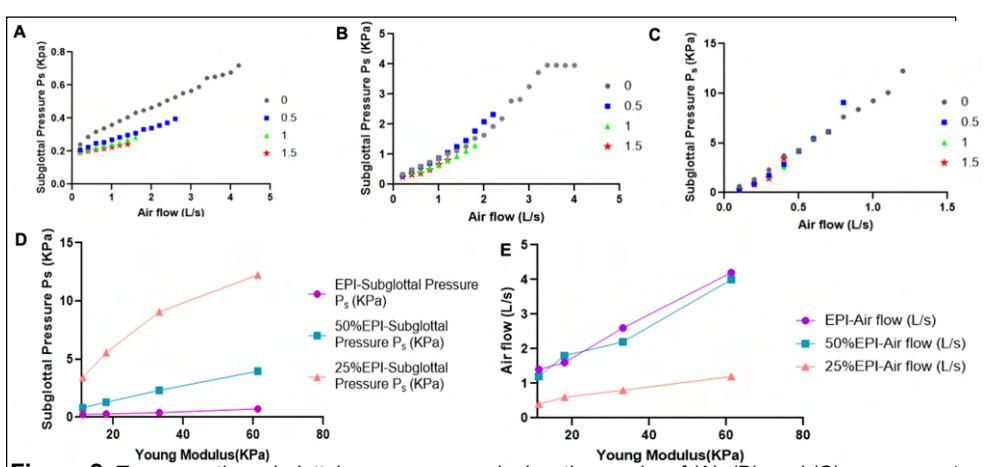


Figure 2. To ensure the subglottal pressure on each size, the graphs of (A), (B), and (C) are 100%, 50%, and 25% of the EPI size, respectively, at different ratios of Ecoflex:thinner ratios; (D) and (E) represent the subglottal pressure and airflow versus the stiffness of the fold (kPa).

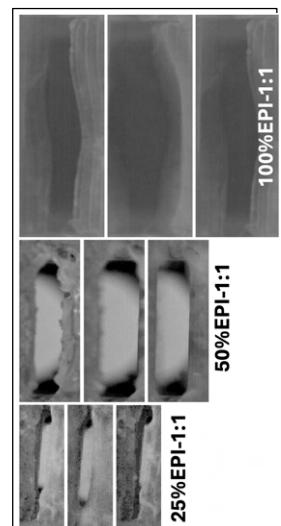


Figure 3. Selected images to show vibration patterns of 100%, 50%, and 25% of the EPI size, respectively, in a fixed airflow (1.6 L/s).

Simulation of flow induced vibration of the vocal folds using the Lattice Boltzmann method

Stefan Kniesburges¹, Bogac Tur¹, Adrian Kummerländer², Mathias J. Krause^{2,3}

¹ Phoniatrics & Pediatric Audiology, University Hospital Erlangen, Erlangen, Germany

² Institute for Applied and Numerical Mathematics, Karlsruhe Institute of Technology, Karlsruhe, Germany

³ Institute for Mechanical Process Engineering and Mechanics, Karlsruhe Institute of Technology, Karlsruhe, Germany

Keywords: *Fluid-structure interaction, vocal folds' vibration, Lattice Boltzmann method*

Abstract

Introduction & rationale:

Continuum-based computational larynx models for simulating the fluid-structure interaction of the vocal folds' vibration enable the simulation of fluid flow around the vocal folds and the response of their deformation. In general, combinations of Finite Volume, Finite Difference and Finite Elements methods are applied to model the physical process. The large drawback of this approach is the extremely long simulation time to simulate the 3D fluid-structurally coupled process, normally in the range of hours to few days for simulating one oscillation cycle. The reasons are the occurring turbulence within the supraglottal flow field and the communication between flow and structural deformation solver of the tissue model.

Nevertheless, the acquired information from those models is very valuable because they include all relevant information of the aero-acoustic sound generation in the larynx. Thus, its relevance for clinical diagnostics and therapy is concluded to be very high.

Objectives:

In this study, the laryngeal flow is simulated with a CFD solver that is based on the Lattice Boltzmann method (LBM) and combined with a lumped-mass model involving 6 masses for each vocal fold. The boundary and geometric conditions are selected in accordance to an experimental silicone model. The aim is to reproduce valid flow-induced oscillation of the vocal folds numerically including complete glottis closure with an acceptably small amount of simulation time.

Methods:

A homogenized LBM (Krause et al., 2021) based on the D3Q19 velocity discretization is employed to simulate the fluid flow. The vocal folds within the fluid domain are treated as a moving porous medium whose geometry is based on the M5 model. Turbulent effects in the supraglottal region are modeled using a large eddy simulation with the Smagorinsky subgrid-scale model. The computational domain measures $0.1 \times 0.01 \times 0.015 \text{ m}^3$ (in the x,y and z directions) and is discretized into approximately two million lattice cells. A pressure inlet, a velocity outlet, and no-slip boundary conditions on the walls are prescribed. The maximum outlet velocity is set to 1 m/s, and a time step of $8.63547 \times 10^{-7} \text{ s}$ is used for the fluid solver.

The vocal folds' structural response is represented by the six-mass model implemented in Python, following the approach of Schwarz et al. (2008). Each vocal fold is composed of six discrete masses that account for various forces: (1) an anchor force with spring-damper-like behavior with a stiffness that varies nonlinearly depending on mass position, (2) vertical and (3) longitudinal linear spring forces between adjacent masses, (4) collision forces to prevent mass overlap, and (5) the driving force derived from the subglottal pressure. The governing ordinary differential equations (ODEs) are integrated using a fourth-order Runge–Kutta method.

A partitioned fluid-structure coupling is realized through the preCICE library (Bungartz et al., 2016). During each fluid time step, velocity and pressure fields are computed via the fluid solver. The local pressure from a designated lattice cell beneath the vocal folds is then passed to the 6MM solver, providing the time-varying subglottal pressure (driving force). Next, the 6MM solver calculates the updated mass positions, which are transferred back to the fluid solver to update the vocal fold geometry. The fluid–solid exchange is carried out every 16th fluid timestep representing a time interval of $1.38 \times 10^{-5} \text{ s}$ which is a result of a coupling timestep independence study.

All simulations are carried out on a local workstation equipped with a medium-performance NVIDIA GeForce RTX 2080 Ti GPU (Nvidia, Inc., Santa Clara, CA).

Results and Discussion:

The two-way coupled fluid–structure simulations show a vocal fold oscillation with a fundamental frequency of approximately 250 Hz. The open quotient and closing quotient are computed to be OQ=0.63 and CIQ=0.30, respectively, indicating a sufficiently prolonged glottal open phase while still achieving complete glottis closure (glottal gap index

GGI=0). Visual inspection of the computed vocal fold deformation reveals a clear mucosal wave propagation, consistent with physiological patterns of oscillation and a maximum medial amplitude of 3.4 mm. Throughout one oscillation cycle, which is shown in Figure 1 for the glottis and the flow field, the vocal folds clearly exhibit a convergent shape during opening and a divergent shape during closing. This shift in glottal geometry markedly influences the airflow patterns and contributes to the characteristic vortex formations downstream. For the chosen coupling timestep between the fluid solver and the six-mass model (structure), the mean simulation time for one oscillation cycle took approx. 8 min having a performance of approx. 50 mio LUPs (Lattice Updates Per sec). The vast majority of this time is spent in the communication between LBM and structure solver - eliminating this in a future version taking full advantage of fluid-only performance of 3 billion LUPs, the computation time for one oscillation cycle will be reduced to approx. 12 s.

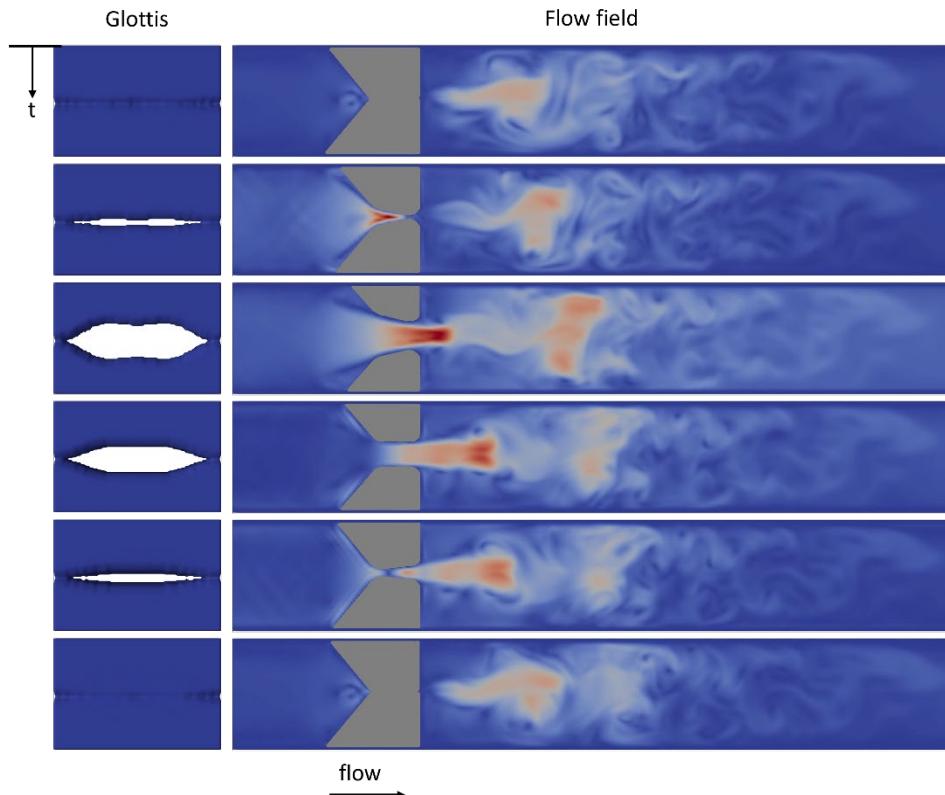


Figure 1: Sequential glottal dynamics and resulting flow field over one oscillation cycle. The left sequence shows the glottis opening over one cycle. The right sequence shows the vocal folds (grey), and the flow field in the supraglottal region. The flow is directed from left to right, and the parameter t represents the time.

Conclusions:

In summary, this dimension reduced approach, integrating a Lattice Boltzmann solver with the six-mass model of the vocal folds, successfully reproduces realistic vocal fold oscillations—including full glottal closure—within a reasonable computational timeframe on a state-of-the-art workstation. Assuming the performance of large GPU clusters, the computational timeframe will even become much shorter as estimated, enabling to analyze detailed information about the laryngeal airflow patterns and tissue deformation for clinical diagnostics and phonation research with large case studies.

References:

- Bungartz, H.-J., Lindner, F., Gatzhammer, B., Mehl, M., Scheufele, K., Shukaev, A., & Uekermann, B. (2016). preCICE – A fully parallel library for multi-physics surface coupling. *Computers & Fluids*, 141, 250–258. <https://doi.org/10.1016/j.compfluid.2016.04.003>
- Mathias J. Krause, M. J., Kummerländer, A., Avis, S.J., Kusumaatmaja, H., Dapelo, D., Klemens, F., Gaedtke, M., Hafen, N., Mink, A., Trunk, R., Marquardt, J.E., Maier, M.-L., Haussmann, M., Simonis, S. (2021). OpenLB—Open source lattice Boltzmann code. *Computers & Mathematics with Applications*, 81, 258–288. <https://doi.org/10.1016/j.camwa.2020.04.033>
- Schwarz, R., Döllinger, M., Wurzbacher, T., Eysholdt, U., & Lohscheller, J. (2008). Spatio-temporal quantification of vocal fold vibrations using high-speed videoendoscopy and a biomechanical model. *The Journal of the Acoustical Society of America*, 123(5), 2717–2732. <https://doi.org/10.1121/1.2902167>

A feasibility study of subharmonic voice detection by deep learning

Takeshi Ikuma¹, Brad Story³, Andrew McWhorter¹, Melda Kunduk^{1,2}

¹Department of Otolaryngology--Head and Neck Surgery, Louisiana State University Health Sciences Center, New Orleans, LA, USA

²Department of Communication Disorders, Louisiana State University, Baton Rouge, LA, USA

³Department of Speech, Language, Hearing Sciences, University of Arizona, Tucson, AZ, USA

Keywords: Disordered voice, Machine learning, Voice signal analysis

Abstract

Introduction & Rationale:

Subharmonic phonation (Behrman et al., 1998; Cavalli & Hirson, 1999; Ikuma et al., 2023) is associated with vocal pathology, especially those which alter the physical characteristics of the vocal folds such as vocal fold lesions and vocal fold paralysis. Not only subharmonic voice is known to cause a rough perceptual quality (Omori et al., 1997), its nearly periodic characteristic may also cause voice analysis algorithms and tools to under-report the voice severity because normal voice is also nearly periodic. As such, a numerical solution to detect the presence of the subharmonics could improve the accuracy of the acoustic analysis of pathological voice.

Voice signals encompassing the subharmonic behaviors are categorized as type 2 in the voice signal typing system (Titze, 1994). Along with defining the signal types, Titze (1994) recommended analyzing type 2 signals by visualization tools and subjective inspection and discouraged the use of perturbation measures citing their unreliability. Altered periodicity of subharmonic signals is a cause of unreliable voice measures. Specialized analysis methods for subharmonics are also still premature. The NSH (number of subharmonics) parameter of KayPENTAX CSL/MDVP software reportedly has poor accuracies (Cavalli & Hirson, 1999). Sun (2000) introduced the concept of the subharmonic-to-harmonic ratio (SHR) and used it to estimate the fundamental frequency of voice signals with a possible presence of period-doubling subharmonics. Sun's method, however, is not the ideal solution to compute the SHR as there is a conundrum of trying to measure the SHR without knowing the presence of the subharmonics. The method is also limited to detect period-doubling subharmonics while higher-period subharmonics are also prevalent in pathological voice (Cavalli & Hirson, 1999; Ikuma et al., 2023).

Deep learning (DL) is a potential solution to detect the presence of subharmonics in voice signals. This idea is motivated by the DL fundamental frequency estimators (Ardaillon & Roebel, 2019; Kim et al., 2018). They outperformed the conventional estimators and demonstrated their robustness against the presence of subharmonics (Ikuma et al., 2024). If a DL network can be trained to ignore subharmonics, a similar network could likely be trained to find subharmonics.

Objectives:

The primary property of the subharmonics is their period, i.e., the number of glottic cycles with varying magnitude, duration, or shape before repeating. Thus, the detection of subharmonics is synonymous to estimation of the subharmonic period M , noting that $M = 1$ naturally lend itself to the normal phonation. Accordingly, the objective of the study was to assess the feasibility of classifying the subharmonic period M using a fully convolutional neural network (FCN) (a type of DL network) and numerically synthesized acoustic signals as the training data.

The FCN under study is shown in Fig. 1. This network takes a running voice signal x_n at 8000 samples/second and produces the probabilities $\{P_{M,k} : M = 1, 2, \dots, 4\}$ of the k th snapshot of 401 input samples (50.1 ms) consisting of subharmonic period M , up to period quadrupling. The output $P_{M,k}$ is produced at the interval of 2 ms. The estimated subharmonic period of the k th snapshot is then selected by $\hat{M}_k = \operatorname{argmax}_M P_{M,k}$.

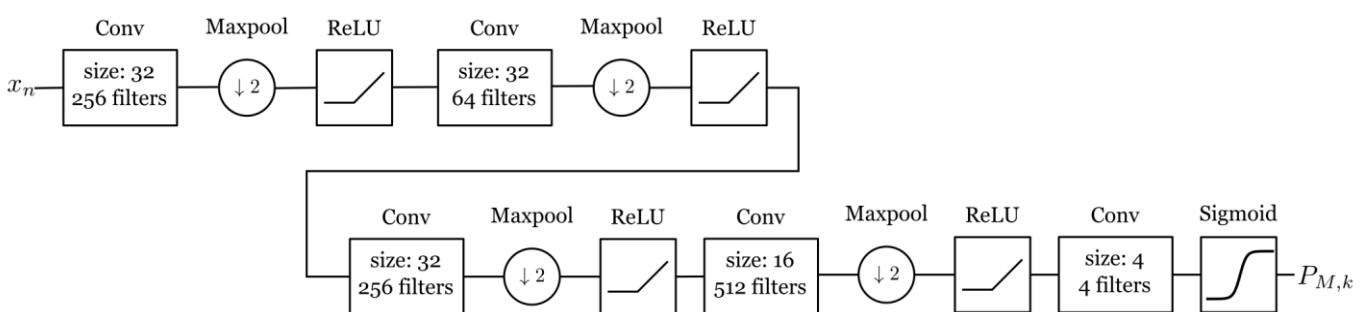


Fig. 1: Fully convolutional neural network under study.

The training dataset was generated by Monte Carlo simulation with the kinematic vocal fold model (Titze, 1984) coupled with the wave-reflection vocal tract model (Story, 1995). The subharmonic vocal fold vibration was programmed to comprise both amplitude and frequency modulations. The subglottal and supraglottal vocal tracts

were modeled as leaky cylinders, connected to a constant lung pressure source and a circular piston in an infinite baffle as the lip radiator. Simulation also used a modified Klatt aspiration noise model (Klatt & Klatt, 1990) to inject colored Gaussian noise at the glottis. The simulation was controlled by 22 parameters, and all except for M were randomly chosen from independent uniform distribution for each run.

Results:

The total of 40000 1-second signal segments (10000 for each M) were generated for the training dataset. Each segment yielded 478 snapshots, each with the same simulation parameters but with different phases and aspiration noise. The dataset was split into 80% training and 20% validation. The FCN was trained over 20 epochs, and the network model with the best classification accuracy of the validation dataset was chosen for the evaluation.

The evaluation dataset of 4000 signal segments were separately generated, including the computation of their SHRs. The subharmonic detection rate was 97.60% with the false alarm rate of 0.45%. Additional results are shown on the confusion matrix in Fig. 2(a). The SHR of the generated subharmonics are shown in Fig. 2(b) and the subharmonic period estimation accuracy dependence on the SHR in Fig. 2(c).

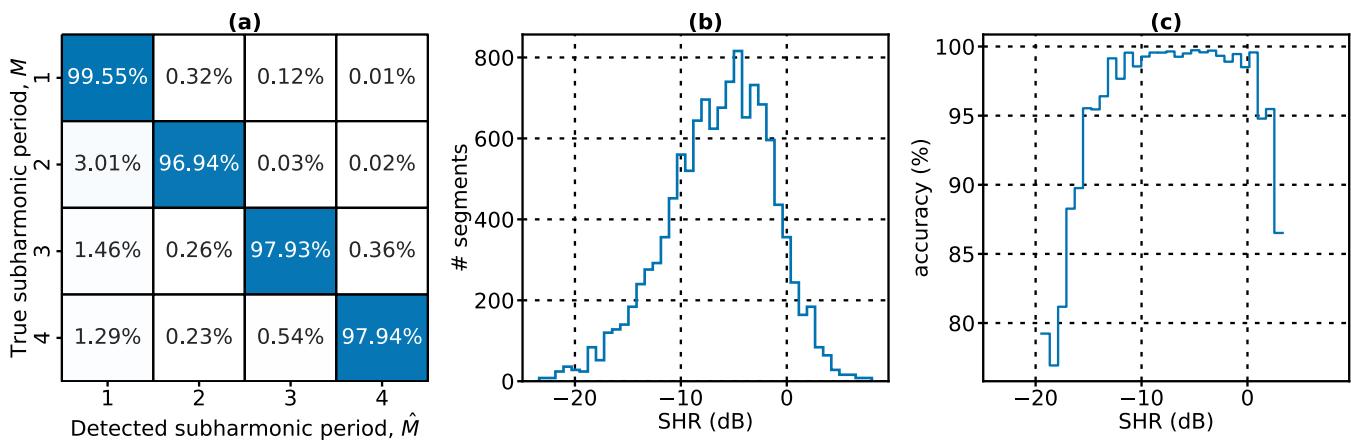


Fig. 2: Estimation performance: (a) confusion matrix, (b) SHR distribution ($M > 1$ only), (c) classification accuracy as a function of SHR (bins with at least 12 segments).

Conclusions:

The synthetic evaluation indicates that the FCN is highly capable of correctly identifying the subharmonic periods, including the normal voice ($M = 1$). Testing with the pathological voice recordings (to be presented) indicates that the FCN is already demonstrating its potential despite being trained with crude vocal tract model and arbitrarily selected parameter ranges. Refining the training dataset will likely yield a dependable subharmonic detector.

References

- Ardaillon, L., & Roebel, A. (2019). Fully-convolutional network for pitch estimation of speech signals. *Interspeech 2019*, 2005–2009. <https://doi.org/10.21437/Interspeech.2019-2815>
- Behrman, A., Agresti, C. J., Blumstein, E., & Lee, N. (1998). Microphone and electroglottographic data from dysphonic patients: Type 1, 2 and 3 signals. *J. Voice*, 12(2), 249–260. [https://doi.org/10.1016/S0892-1997\(98\)80045-3](https://doi.org/10.1016/S0892-1997(98)80045-3)
- Cavalli, L., & Hirson, A. (1999). Diplophonia reappraised. *J. Voice*, 13(4), 542–556. [https://doi.org/10.1016/S0892-1997\(99\)80009-5](https://doi.org/10.1016/S0892-1997(99)80009-5)
- Ikuma, T., McWhorter, A. J., Adkins, L., & Kunduk, M. (2023). Investigation of vocal bifurcations and voice patterns induced by asymmetry of pathological vocal folds. *J. Speech. Lang. Hear. Res.*, 66(1), 48–60. https://doi.org/10.1044/2022_JSLHR-21-00499
- Ikuma, T., McWhorter, A. J., & Kunduk, M. (2024). Evaluation of machine-learning pitch estimation algorithms. *13th ICVPB*, 28–29.
- Kim, J. W., Salamon, J., Li, P., & Bello, J. P. (2018). Crepe: A convolutional representation for pitch estimation. *IEEE ICASSP 2018*, 161–165. <https://doi.org/10.1109/ICASSP.2018.8461329>
- Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *J. Acoust. Soc. Am.*, 87(2), 820–857. <https://doi.org/10.1121/1.398894>
- Omori, K., Kojima, H., Kakani, R., Slavit, D. H., & Blaugrund, S. M. (1997). Acoustic characteristics of rough voice: Subharmonics. *J. Voice*, 11(1), 40–47. [https://doi.org/10.1016/S0892-1997\(97\)80022-7](https://doi.org/10.1016/S0892-1997(97)80022-7)
- Story, B. (1995). *Physiologically-Based Speech Simulation Using an Enhanced Wave-Reflection Model of the Vocal Tract* [PhD thesis]. University of Iowa.
- Sun, X. (2000). A pitch determination algorithm based on subharmonic-to-harmonic ratio. *Proc. 6th ICSLP*, 4, 676–679. <https://doi.org/10.21437/ICSLP.2000-902>
- Titze, I. R. (1984). Parameterization of the glottal area, glottal flow, and vocal fold contact area. *J. Acoust. Soc. Am.*, 75(2), 570–580. <https://doi.org/10.1121/1.390530>
- Titze, I. R. (1994). *Workshop on Acoustic Voice Analysis: Summary Statement*. National Center for Voice and Speech.

PODIUM SESSION 9

Acoustic-Aerodynamic Vocal Efficiency Metrics Identify Changes in Vocal Function Following Voice Therapy for Phonotraumatic Vocal Hyperfunction Subgroups

Emma C. Willis^{1,2}, Annie B. Fox¹, Jarrad Van Stan^{1,3}, Jameson Cooper², Ben Kevelson²,

Robert E. Hillman^{1,3}, Daryush D. Mehta^{1,3}

¹ School of Health & Rehabilitation Sciences, MGH Institute of Health Professions, Boston, MA, USA

² Center for Laryngeal Surgery and Voice Rehabilitation, Massachusetts General Hospital, Boston, MA, USA

³ Harvard Medical School, Boston, MA, USA

Keywords: *Vocal efficiency; Voice assessment; Vocal hyperfunction*

Abstract

Introduction & rationale: In the United States, 12.2% of the population experiences a voice disorder annually (Naunheim et al., 2024), with a 30% lifetime prevalence (Roy et al., 2013). As such, many individuals will be impacted by a voice disorder of some kind during their life. Vocal hyperfunction is considered an etiological component in the most commonly occurring types of voice disorders (Hillman et al., 2020). It can be defined as “excessive peri-laryngeal musculoskeletal activity during phonation” (Oates & Winkworth, 2008). In the vocal hyperfunction framework (Hillman et al., 2020), phonotraumatic vocal hyperfunction (PVH) is associated with obvious signs of tissue trauma on the true vocal folds. Treatment options for PVH include surgical intervention, voice therapy, or a combination of the two. In many mild and moderate cases, voice therapy as the only treatment is common for patients with PVH.

The presence of PVH can impact vocal fold biomechanics including glottic closure which can result in atypical values for mean airflow, sound pressure level, and subglottal pressure. These changes can cause reduced vocal efficiency. Vocal efficiency is a quantitative method to evaluate how efficiently aerodynamic energy is converted to audible sound (Titze, 1992). Depending on the severity of the hyperfunction and/or lesions, the changes in these measures may be more or less pronounced. Previous work (Toles et al., 2022) has already investigated how vocal efficiency changes pre/post-surgical intervention in individuals diagnosed with PVH. However, to the authors’ knowledge, no work has been done to quantify changes in vocal efficiency pre/post-voice therapy as the sole intervention for PVH. The most common methods of calculating vocal efficiency have included the (1) traditional voicing efficiency equation, (2) ratio of sound pressure level (SPL) and subglottal pressure (Ps) as SPL/Ps, (3) ratio of SPL and mean airflow, and (4) the ratio of SPL and aerodynamic power (AP; product of mean airflow and subglottal pressure) (SPL/AP). Bullock et al. (2023) had findings suggestive that SPL/Airflow may be more strongly associated with treatment effects than other vocal efficiency metrics. They also found that detection of changes in vocal efficiency in patients with unilateral vocal fold paralysis pre/post treatment could be influenced by the amount and direction of change in aerodynamic measurement. This suggests that there may be subgroups of individuals that behave similarly (e.g., people who tended to have high average airflow versus people who tended to have low average airflow) despite sharing the same overall diagnosis. This study investigates all four vocal efficiency measures to assess if certain voicing efficiency metrics are more strongly associated with voice therapy treatment effects in patients with PVH.

Objectives: This study aimed to 1) quantify how vocal efficiency changes pre/post voice therapy in patients with PVH, 2) compare how therapy changes relate to normative data, and 3) evaluate how four different vocal efficiency metrics compare in the detection of treatment effects.

Methods: Acoustic and aerodynamic measures at soft, comfortable, and loud volumes were collected and analyzed from 121 female participants with a diagnosis of PVH and 203 vocally healthy female control participants between the ages of 18–65 years. Diagnoses were based on a comprehensive team evaluation (laryngologist and speech-language pathologist) at the Massachusetts General Hospital Center for Laryngeal Surgery and Voice Rehabilitation. The vocally healthy status of all matched-control participants was verified via interview, auditory-perceptual evaluation by a speech-language pathologist, and a laryngeal stroboscopic examination. Acoustic and aerodynamic measures were collected at two time points (baseline and follow up) for all participants. The baseline and follow up time points were, respectively, before and after voice therapy for participants with PVH, and initial visit and 6-month follow-up visit for control participants. Each participant was instructed to produce monotonic consonant-vowel strings (/pa:pa:pa:pa:pa/) in three loudness conditions (soft, comfortable, loud). Multilevel models were utilized for analysis of the data on both stratified and unstratified groups. Since loudness condition in initial analysis was found to have a significant impact on vocal efficiency in the unstratified group, subsequent analyses were done on each loudness condition separately. Stratification for vocal efficiency level for subgrouped analyses was based on percentiles derived from the control group (where >0.66 is high, between $0.33-0.66$ is middle, and <0.33 is low) for each loudness condition. Marginal R² values were computed using the method proposed by Nakagawa & Schielzeth (2013) for comparison of the four vocal efficiency metrics.

Results: Results showed several significant effects of treatment (Figure 1). In the comfortable loudness condition, traditional voicing efficiency and SPL/Ps detected significant treatment effects for the high vocal efficiency subgroup ($d = -0.76$ and -1.02 , respectively) and low vocal efficiency subgroup ($d = 0.74$ and 0.90 , respectively). SPL/Airflow and SPL/AP detected

significant treatment effects for the high vocal efficiency subgroup only ($d = -1.39$ and -1.99 , respectively). Results were generally consistent in the soft and loud conditions for SPL/Ps and SPL/AP, with some differences for traditional voicing efficiency seen in both the soft and loud conditions and just in the soft condition for SPL/Airflow. Marginal R^2 values were extracted from the patient-only models to evaluate the ability to detect treatment effects using the four vocal efficiency metrics. Initial analysis of marginal R^2 values in the comfortable loudness condition revealed that the model for SPL/Ps accounted for the most fixed effects ($R^2_m = 55.82\%$), followed by traditional vocal efficiency ($R^2_m = 54.20\%$), SPL/AP ($R^2_m = 51.47\%$), and SPL/Airflow ($R^2_m = 44.80\%$).

Conclusions: Pre-therapy vocal efficiency values for PVH patients range from atypically high to atypically low, and statistically significant post-therapy treatment effects were only revealed when patients are stratified/subgrouped based on the pre-therapy vocal efficiency measures (high, middle, low). Following treatment, vocal efficiency tended to normalize, or show evidence of moving in the direction of normalization, depending on which vocal efficiency metric was used. All measures demonstrated large effect sizes and marginal R^2 values, however traditional vocal efficiency and SPL/Ps performed best in detecting treatment effects in both the high and low patient subgroups with large effect sizes and the largest marginal R^2 values. These results support the continued investigation/development of vocal efficiency for clinical voice assessment.

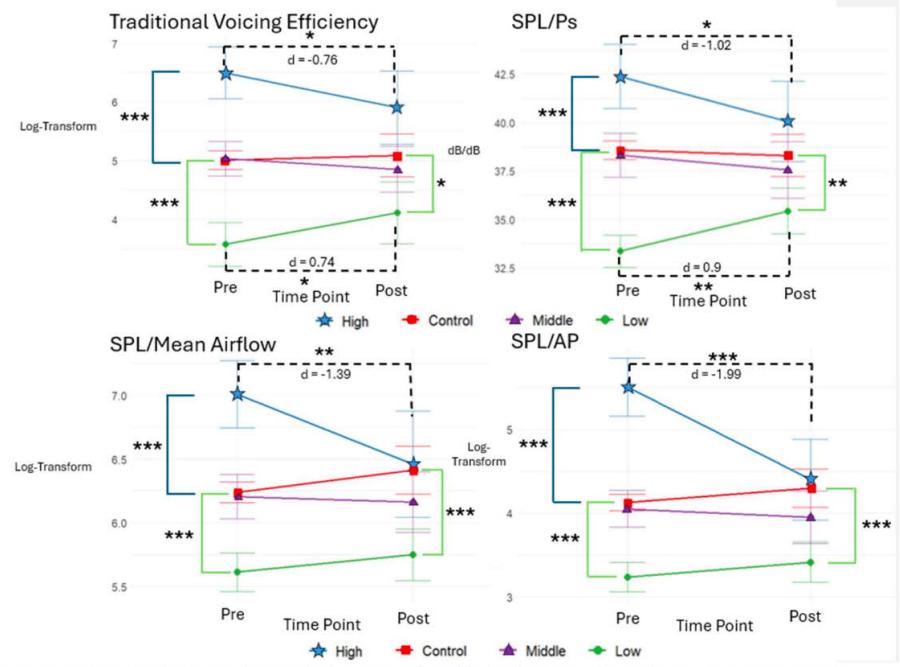


Figure 1: Patients with PVH and controls during comfortable loudness condition demonstrating treatment effects for high and low vocal efficiency subgroups.

Following treatment, vocal efficiency tended to normalize, or show evidence of moving in the direction of normalization, depending on which vocal efficiency metric was used. All measures demonstrated large effect sizes and marginal R^2 values, however traditional vocal efficiency and SPL/Ps performed best in detecting treatment effects in both the high and low patient subgroups with large effect sizes and the largest marginal R^2 values. These results support the continued investigation/development of vocal efficiency for clinical voice assessment.

References:

- Bullock, L., Toles, L. E., Hillman, R. E., & Mehta, D. D. (2023). Acoustic-Aerodynamic Voice Outcome Ratios Identify Changes in Vocal Function Following Vocal Fold Medialization for Unilateral Vocal Fold Paralysis. *Journal of Voice : Official Journal of the Voice Foundation*, S0892-1997(23)00104-2. <https://doi.org/10.1016/j.jvoice.2023.03.011>
- 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. https://doi.org/10.1044/2020_AJSLP-20-00104
- Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, 4(2), 133-142.
- Naunheim, M. R., DeVore, E. K., Huston, M. N., Song, P. C., Franco Jr, R. A., & Bhattacharyya, N. (2024). Increasing Prevalence of Voice Disorders in the USA: Updates in the COVID Era. *The Laryngoscope*, 134(8), 3713–3718. <https://doi.org/10.1002/lary.31409>
- Nelson R., Barkmeier-Kraemer J., Eadie T., Sivasankar M.P., Mehta D., Paul D., & Hillman R. (2013). Evidence-Based Clinical Voice Assessment: A Systematic Review. *American Journal of Speech-Language Pathology*, 22(2), 212–226. [https://doi.org/10.1044/1058-0360\(2012/12-0014\)](https://doi.org/10.1044/1058-0360(2012/12-0014))
- Oates, J., & Winkworth, A. (2008). Current knowledge, controversies and future directions in hyperfunctional voice disorders. *International Journal of Speech-Language Pathology*, 10(4), 267–277. <https://doi.org/10.1080/17549500802140153>
- Titze, I. R. (1992). Vocal efficiency. *Journal of Voice*, 6(2), 135–138. [https://doi.org/10.1016/S0892-1997\(05\)80127-4](https://doi.org/10.1016/S0892-1997(05)80127-4)
- Toles L.E., Seidman A.Y., Hillman R.E., & Mehta D.D. (2022). Clinical Utility of the Ratio of Sound Pressure Level to Subglottal Pressure in Patients Surgically Treated for Phonotraumatic Vocal Fold Lesions. *Journal of Speech, Language, and Hearing Research*, 65(8), 2778–2788. https://doi.org/10.1044/2022_JSLHR-21-00658

Respiratory kinematics in aging singers and non-singers: relationship to vocal effort

Clotilde Marc¹, Jessica E. Huber², Maude Desjardins³

¹Speech therapy Department, Faculty of Health, University of Rouen Normandy, Rouen, France

² Department of Communicative Disorders and Sciences, University at Buffalo, Buffalo, NY, United States ³ School of Rehabilitation Sciences, Université Laval, Quebec City, Quebec, Canada

Keywords: respiratory kinematics, vocal effort, singing

Abstract

Introduction & rationale:

Previous studies have identified relationships between respiratory kinematics and laryngeal function during speech. Phonating at lower lung volumes promotes a higher laryngeal position, a greater closed quotient, more supraglottic compression, and lower subglottal pressures than phonating at higher lung volumes¹⁻³. Moreover, lower lung volumes during demanding tasks have been reported in speakers experiencing vocal fatigue when compared to those without vocal complaints^{4, 5}. These findings highlight the importance of satisfactory respiratory planning to achieve speech and vocal goals. In aging speakers, proper respiratory planning is especially important considering decreases in lung elastic recoil – associated with a loss of natural positive pressures in the lungs – and reduced glottal closure promoting rapid air pressure loss. Older adults frequently experience speech-related fatigue, which can restrict social participation^{6, 7}, and optimal use of the respiratory system may help alleviate these symptoms. However, factors promoting optimal planning for speech, i.e., respiratory kinematics resulting in the least vocal effort, are not known.

Similar to other complex coordinated actions, respiratory planning is thought to rely on the adaptation of motor programs through the comparison of sensory feedback to anticipated sensory targets^{8, 9}. Therefore, we hypothesized that respiratory planning is an experience-dependent skill that may be impacted by singing practice, analogous to other aspects of vocal control such as pitch accuracy¹⁰.

Objectives:

The objectives of this study were: (1) To investigate the impact of singing status on respiratory kinematics for speech in middle-aged and older speakers. We hypothesized that singing experience would promote more precise respiratory planning, i.e., that singers would exhibit smaller differences between lung volume planned at speech initiation and lung volume exerted. (2) To assess the relationship between key measures of respiratory kinematics and perceived vocal effort in middle-aged and older singers and non-singers. We hypothesized that phonating at lower lung volumes would be associated with increased vocal effort in speakers with and without singing experience.

Methods:

Participants and tasks

The study sample comprised twenty-nine middle-aged and older participants without vocal complaints (mean: 68.45 years; range: 50-91 years) divided into 18 amateur singers (recruited from local choirs) and 11 non-singers. Participants read sentences of various lengths (3-72 syllables) in a comfortable voice and in a loud voice, while wearing inductive respiratory bands and speaking into a headset microphone. Loudness conditions were randomized, and after each condition participants were instructed to indicate their perceived vocal effort on a modified Borg scale (rating of 0-10, with higher ratings reflecting more perceived effort).

Measurements and analyses

Prior to the speech task, movements of the chest wall (rib cage and abdomen) were calibrated to actual lung volumes using a spirometer and custom Matlab script, allowing the estimation of lung volume during speech tasks. At this stage, data or partial data from three participants had to be excluded from the analyses due to unsatisfactory calibration. Estimated lung volume initiation relative to end-expiratory level (LVI), lung volume termination relative to end-expiratory level (LVT), and lung volume excursion (LVE) were then extracted from the speech tasks and expressed as percent of vital capacity (%VC). A deviation score was computed using the following formula: $\log_{10}(LVI/LVE)$ to characterize discrepancies between lung volume planned and lung volume exerted during speech. Mann-Whitney U tests were conducted to assess differences in respiratory and voice measures between singers and non-singers. To investigate relationships between respiratory measures and vocal effort, Spearman correlation coefficients were computed in each group (singers and non-singers) and for the whole sample. Considering the small sample size and preliminary nature of this study, we considered results to be significant** ($p<0.05$) or marginally significant* ($p<0.1$).

Results:

LVI, LVT, LVE, and deviation score were all descriptively higher in the group of non-singers, but the differences did not reach statistical significance except for deviation score which was marginally significant (see Table 1). There were no group differences in terms of vocal effort and vocal intensity.

Table 1. Respiratory and voice measures in singers and non-singers

Measure	Singers	Non-singers	p-value
LVI (%VC)	34.84	40.66	0.15
LVT (%VC)	-0.84	2.44	0.26
LVE (%VC)	36.16	39.17	0.58
Deviation score	0.02	0.09	0.084*
Vocal effort (0-10)	2.13	2.17	0.95
Vocal intensity (db)	73.5	71.71	0.56

LVI was not correlated with perceived vocal effort during the comfortable or loud speech production in either group. In non-singers, a marginally significant inverse and moderate correlation was observed between LVT and perceived vocal effort in the loud voice condition ($r=-0.61$, $p=0.08$). In singers, a significant moderate correlation was observed between LVE and vocal effort in the comfortable voice condition ($r=0.52$, $p=0.033$). No significant relationships were found between respiratory measures and vocal effort when considering the whole sample (singers and non-singers combined).

Conclusions:

Findings suggest that singing experience may impact respiratory planning strategies for speech. Non-singers tended to breathe in more air than needed when compared to singers, whose deviation scores were lower. This is consistent with the fact that non-singers reported greater vocal effort when terminating loud speech at lower lung volumes, i.e., closer to or below EEL. Higher deviations scores may therefore represent an adaptive strategy in non-singers (rather than imprecise planning), suggesting they plan to terminate speech above EEL. This compensation may not be as important in singers. These results, combined with the positive correlation found between LVE and vocal effort in singers, suggest a possible tendency for respiratory parsimony in singers that warrants further investigation.

References:

- Iwarsson J, Sundberg J. Effects of lung volume on vertical larynx position during phonation. *Journal of Voice*. 1998;12(2):159-65.
- Iwarsson J, Thomasson M, Sundberg J. Effects of lung volume on the glottal voice source. *Journal of Voice*. 1998;12(4):424-33.
- Milstein CF. Laryngeal function associated with changes in lung volume during voice and speech production in normal speaking women. [Doctoral dissertation, The University of Arizona]1999.
- Lowell SY, Barkmeier-Kraemer JM, Hoit JD, Story BH. Respiratory and laryngeal function during spontaneous speaking in teachers with voice disorders. *Journal of speech, language, and hearing research*. 2008;51(2):333-49.
- Fujiki RB, Huber JE, Sivasankar MP. The effects of vocal exertion on lung volume measurements and acoustics in speakers reporting high and low vocal fatigue. *PloS one*. 2022;17(5):e0268324.
- Lindström E, Öhlund Wistbacka G, Lötvall A, Rydell R, Lyberg Ålander V. How older adults relate to their own voices: a qualitative study of subjective experiences of the aging voice. *Logopedics Phoniatrics Vocology*. 2023;48(4):163-71.
- Porcaro CK, Singer C, Djokic B, Danesh AA, Tappen R, Engstrom G. Perceived voice disorders in older adults and impact on social interactions. *Perspectives of the ASHA Special Interest Groups*. 2021;6(1):143-50.
- Heller Murray ES, Michener CM, Enflo L, Cler GJ, Stepp CE. The Impact of Glottal Configuration on Speech Breathing. *Journal of voice*. 2018;32(4):420-7.
- Desjardins M, Jomphe V, Lagadec-Gaulin L, Cohen M, Verdolini Abbott K. Influence of Sensory Monitoring on Speech Breathing Planning Processes: An Exploratory Study in Aging Speakers Reporting Dyspnea. *Journal of Speech, Language, and Hearing Research*. 2024;67(8):2483-98.
- Zarate JM, Zatorre RJ. Experience-dependent neural substrates involved in vocal pitch regulation during singing. *NeuroImage*. 2008;40(4):1871-87.

MODELING LARYNGEAL MOTOR CONTROL WITH SOMATOSENSORY FEEDBACK FOR PREPHONATORY VOCAL POSTURE

Nicolás F. Quinteros^{1,2}, Jesús A. Parra², Gabriel A. Alzamendi^{3,4}, Matías Zañartu^{1,2}

¹ Department of Electronic Engineering, Universidad Técnica Federico Santa María, Valparaíso, Chile

² Advanced Center for Electrical and Electronic Engineering, Universidad Técnica Federico Santa María, Valparaíso, Chile

³ Institute for Research and Development on Bioengineering and Bioinformatics (IBB), CONICET-UNER, Oro Verde, Entre Ríos 3100, Argentina

⁴ Facultad de Ingeniería, Universidad Nacional de Entre Ríos, Entre Ríos, Argentina

Keywords: Somatosensory Feedback, Prephonatory Vocal Posture, Laryngeal Motor Control

Introduction & rationale: The adjustment of vocal fold posture from a resting respiratory state to a proper prephonatory posture is a critical aspect of laryngeal motor control, involving the rapid and synergistic coordination of the intrinsic laryngeal muscles (ILM). This transition, occurring before any sound is produced, creates a unique environment in which the somatosensory mechanisms responsible for vocal control can be isolated and examined. Despite its importance, the role of somatosensory feedback in prephonatory posturing remains poorly understood. Recent reviews by Hernández-Morato et al. (2023) and Kent (2024) suggest that somatosensory feedback plays a pivotal role in fine-tuning laryngeal motor control, especially during rapid transitions. Existing neurocomputational models of laryngeal motor control (Weerathunge et al., 2022, Palaparthi et al, 2024) have limited representations of laryngeal somatosensory feedback at the prephonatory stage, which in turn limits their ability to accurately simulate the role of laryngeal proprioceptive and tactile feedback in motor adaptation tasks in this transient phase. By developing a model of laryngeal motor control that integrates somatosensory feedback during the prephonatory phase, insights can be gained into the feedback mechanisms for reaching a phonatory steady state of vocal posture. Our hypothesis posits that a selective set of somatosensory features is essential for achieving and maintaining accurate laryngeal posture prior to phonation, offering potential insights and advances in understanding vocal pathologies.

Objective: To develop a physiologically relevant somatosensory feedback scheme that incorporates prephonatory gestures for a neurocomputational model of laryngeal motor control.

Methods: The triangular body-cover model (TBCM) (Alzamendi et al., 2022), based on the work of Titze and Hunter (2007), was used to represent vocal posture dynamics during the prephonatory phase, simulating the relationship between ILM activations and the resulting glottal configuration. The model inputs consisted of the normalized activations between 0 (no activation) and 1 (full activation) for the five ILM: cricothyroid (CT), thyroarytenoid (TA), lateral cricoarytenoid (LCA), interarytenoid (IA), and posterior cricoarytenoid (PCA). Prephonatory posture is described by four biomechanical features: vocal process coordinates (ξ, ψ), arytenoid rotation angle (θ_a), and vocal fold length (L_g).

Iteratively solving the complete set of differential equations in the TBCM (plant model) for each configuration within a motor control framework is computationally expensive. Here, a deep neural network regressor was employed to approximate the nonlinear relationships between ILM activations and biomechanical outcomes, significantly reducing computational costs while preserving predictive accuracy. Training data were drawn from a total of 161,051 simulations performed that comprehensively map the steady-state laryngeal configurations, corresponding to 0.1 increments for each of the normalized ILM activations, as shown in Figure 1. Feature importance analysis revealed that PCA was crucial for predicting ξ and θ_a , CT strongly influenced ψ and L_g , TA was significant for ψ and L_g , while LCA and IA explained poorly the features.

Somatosensory-feedback control of vocal posture was implemented using an inverse Jacobian method, iteratively adjusting motor commands based on somatosensory errors to refine muscle activations and achieve the target vocal configurations. Given the sensitivity of the control model to initial muscle activations, ILM were initialized at an activation level of 0.1, with a standard deviation of 0.001 across 100 Monte Carlo simulations to assess control consistency. The simulations of the prephonatory vocal posture control model were conducted using ILM activations as inputs and two different sets of proprioceptive somatosensory outputs: one with four features and another with three features, excluding L_g . The rationale behind simulating both conditions was to evaluate whether vocal posture can be effectively defined using only purely proprioceptive features (ξ, ψ , and θ_a) without relying on L_g , which also involves mechanoreceptive feedback. While L_g integrates the influence of all

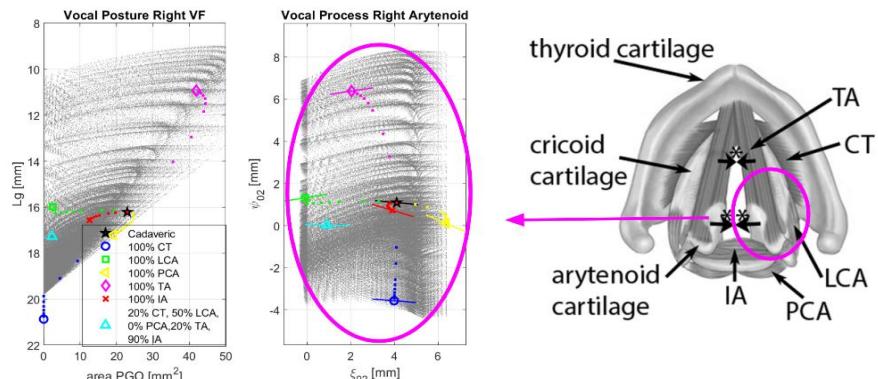


Figure 1: Vocal posture simulations obtained from the TBCM.

muscles, it was excluded to prioritize direct proprioceptive control strategies and assess how the remaining outputs alone could drive stable phonatory postures.

Results: In Figure 2, the left column shows the muscle activation trajectories over time, while the right column presents the evolution of four proprioceptive outputs: (ξ, ψ) , θ_a , and Lg . The system effectively stabilizes these outputs, reaching the target vocal posture with ILM activations converging close to their expected values (0% PCA, 50% LCA, 90% IA, 20% TA, 20% CT). On the other hand, Figure 3 removes Lg as somatosensory information to be controlled, slightly reducing the accuracy of ILM activation estimation and motor control compared to the case with four variables. Although Lg is not used as a control variable but only as a hidden state, the intended vocal process remains well-defined. This highlights that while the inclusion of Lg enhances the precision of achieving target activations, its absence provides a clearer evaluation of the role of purely proprioceptive features in vocal motor control.

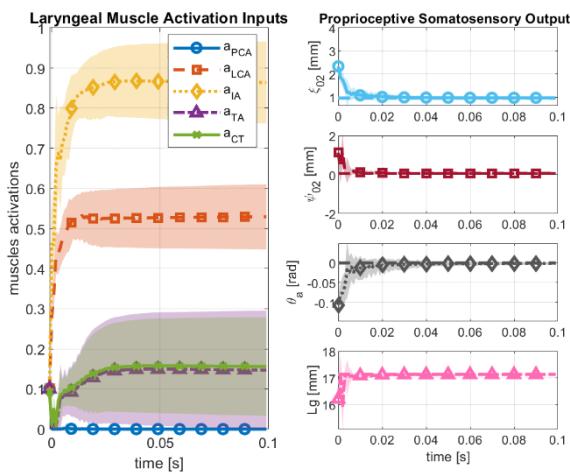


Figure 2: Trajectories of ILM activations to reach a target vocal posture described by four somatosensory features.

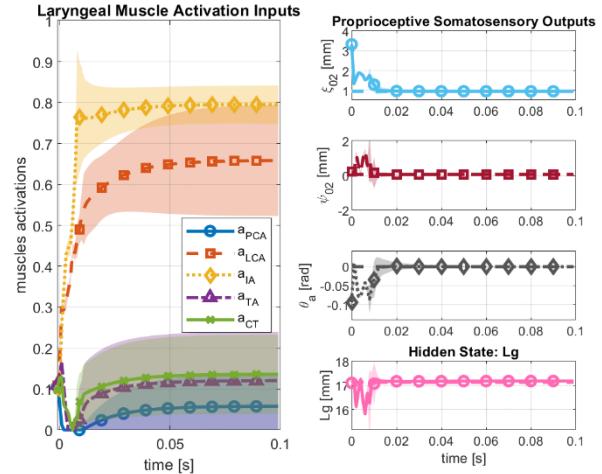


Figure 3: Trajectories of ILM activations to reach a target vocal posture described by three somatosensory features.

Conclusions: This study developed a somatosensory feedback-based model of vocal posture motor control, which provides a novel approach to studying how prephonatory motor control is regulated. Results show that three key proprioceptive biomechanical features are sufficient to achieve and maintain a stable vocal posture, even without explicitly controlling vocal fold elongation. While including one contributes to improved accuracy in targeting muscle activations, excluding one demonstrated that proprioceptive signals alone can also effectively drive prephonatory control. These findings reinforce the idea that proprioceptive feedback plays a critical role in vocal motor control, shaping how the larynx is stabilized prior to phonation.

Future work will focus on refining the model with adaptive learning strategies, similar to the DIVA (Tourville et al., 2011) and LaDIVA (Weerathunge et al, 2022) models, and assessing how these control mechanisms can differ in typical and hyperfunctional voices.

Acknowledgements: This research was supported by the National Institutes of Health (NIH) National Institute on Deafness and Other Communication Disorders grant P50 DC015446, ANID grants FONDECYT 1230828, BASAL AFB240002, and National PhD Scholarship 21240471.

References:

- Alzamendi, G. A., Peterson, S. D., Erath, B. D., Hillman, R. E., & Zañartu, M. (2022). Triangular body-cover model of the vocal folds with coordinated activation of the five intrinsic laryngeal muscles. *The Journal of the Acoustical Society of America*, 151(1), 17-30.
- Hernández-Morato, I., Yu, V. X., & Pitman, M. J. (2023). A review of the peripheral proprioceptive apparatus in the larynx. *Frontiers in Neuroanatomy*, 17, 1114817.
- Kent, R. D. (2024). The Feel of Speech: Multisystem and Polymodal Somatosensation in Speech Production. *Journal of Speech, Language, and Hearing Research*, 67(5), 1424-1460.
- Palaparthi, A., Alluri, R. K., & Titze, I. R. (2024). Deep Learning for Neuromuscular Control of Vocal Source for Voice Production. *Applied Sciences*, 14(2), 769.
- Titze, I. R., & Hunter, E. J. (2007). A two-dimensional biomechanical model of vocal fold posturing. *The Journal of the Acoustical Society of America*, 121(4), 2254-2260.
- 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.
- Weerathunge, H. R., Alzamendi, G. A., Cler, G. J., Guenther, F. H., Stepp, C. E., & Zañartu, M. (2022). LaDIVA: A neurocomputational model providing laryngeal motor control for speech acquisition and production. *PLoS computational biology*, 18(6), e1010159.