ICVPB

Erlangen, Germany

2024

July, 22-27

Programme

Uniklinikum Erlangen

FAU Friedrich-Alexander-Universität Erlangen-Nürnberg

DFG Deutsche Forschungsgemeinschaft
German Research Foundation

13th International Conference on Voice Physiology and Biomechanics

HÖREN SCHENKEN e.V.

Contents

Welcome

It is with great pleasure and honor that we welcome you to Erlangen for the *13th International Conference on Voice Physiology and Biomechanics*. Over the coming days, we will embark on an exciting journey through the latest findings and developments in the field of voice physiology, pathology and biomechanics. This is an opportunity to push the boundaries of our knowledge and learn from each other.

Erlangen, as the host city of this prestigious event, not only offers an inspiring environment but also a rich history and culture that will most hopefully contribute to a setting filled with lots of life and joy. We encourage you, therefore, not only to participate in the conference events but also to explore the beauty and diversity of this city. Come visit the traditional Bavarian "Biergarten" with us and taste some locally brewed beers. Learn something about the historic city of Nuremberg on our guided tour immediately before the Gala Dinner at the Bratwurst Röslein restaurant, that was founded back in 1431. Enjoy a unique atmosphere at our singing evening in the historic beer cellars of Erlangen. Make the most out of your stay.

Our conference brings together some of the brightest minds and leading experts in the field of voice physiology and biomechanics. Through the exchange of ideas, experiences, and research findings, we will contribute to paving the way for future discoveries and innovations. Please actively engage in discussions, ask questions, and share your knowledge with others. We are looking forward to numerous compelling talks and posters on the three days of our conference. Also, we are pleased to offer you a variety of exciting workshops that will provide valuable insights and hands-on experiences in current topics.

As we gather here, it is important to acknowledge that the human voice is not only a scientific phenomenon but also holds cultural and artistic significance. From communication to artistic expression, the voice is an essential part of human life. The ability to talk, communicate and sing is – in our peculiar form – an essential and uniquely human task that affects all aspects of our every-day life. Therefore, we should always consider its complexity and versatility in our thinking.

Importantly we would like to express our heartfelt thanks to the workshop organizers, the Deutsche Forschungsgemeinschaft (DFG) for their generous funding (Grant No. KN 1331/8-1) and supporters of this conference, who have made it possible for us to come together and stand at the forefront of voice research. We also would like to express our sincere gratitude to all participants. Your presence and contributions are essential to the success of this conference. Science and research are a dynamic process, each of us contributes to in a different, solitary way and it's the togetherness which makes it open for different directions and finally expand.

We are proud to welcome you here in Erlangen! And now it's up to you – enjoy the conference!

Michael Döllinger and Stefan Kniesburges

Organization Committee

Prof. Dr. Michael Döllinger
General Chairman
Division of Phoniatrics and Pedaudiology
University Hospital Erlangen
michael.doellinger@uk-erlangen.de
Tel.: +49 (0) 9131 85 33814

PD Dr. Stefan Kniesburges
General Chairman
Division of Phoniatrics and Pedaudiology
University Hospital Erlangen
stefan.kniesburges@uk-erlangen.de
Tel.: +49 (0) 9131 85 32616

M.Sc. Jonas Donhauser
Research Associate Computational Medicine
jonas.donhauser@uk-erlangen.de
Tel.: +49 (0) 9131 85 32607

Ute Katz
Conference Secretary
ute.katz@uk-erlangen.de
Tel.: +49 (0) 9131 85 33814

B.Sc. Ann-Kathrin Kopp
Conference Secretary
ann-kathrin.kopp@uk-erlangen.de
Tel.: +49 (0) 9131 85 33814

We are always happy to help!

About the region

"Bier. Bratwurst. Bayern."

If you ask people outside Bavaria to describe the Nuremberg Metropolitan Region in just three words, you often get at least one of these answers. And people are not wrong. But it has much more to offer, and we would like to encourage you to discover as much of this historic region as possible.

The Nuremberg Metropolitan Region is one of eleven metropolitan regions in Germany and is home to around 3.5 million people. The main cities are Nuremberg, Bamberg, Bayreuth, Fürth and Erlangen, where the conference will take place.

Nuremberg Old Town, Source: schulzfoto/stock.adobe.de

Erlangen – A city of Bavarian clichés

The obvious facts first. Erlangen is a Middle Franconian city in Bavaria, in the south of Germany. While you probably know that Bavaria is most famous for its beer, most people don't know that Erlangen used to be Bavaria's number one export city for beer - even ahead of Munich. 150 years ago, there were 18 breweries in the Huguenot city. The special thing about Erlangen were the beer cellars in a hill aligned to the city called the *Erlanger Burgberg*, which kept the beer cool and fresh even in the hot summer months. This advantage was lost with the invention of refrigeration systems in the 19th century, but the cellars are still in operation today and you can sample traditionally brewed beer in one of the many "beer gardens" along the *Burgberg*. Here you can also try many Franconian specialties such as bratwurst.

"Der Berg ruft!" – engl.: "The mountain is calling!"

Each year during the twelve days before and after Pentecost, Erlangen becomes the scene for the world´s oldest Beer Festival – the *Bergkirchweih*. Originated in 1755, it is even older than the *Oktoberfest*. People of all ages from all around the world gather to enjoy music, cold beer out of stoneware jugs, Franconian specialties and many attractions in front of the historic beer cellars. With around one million visitors a year, it is smaller than the Oktoberfest but therefore rather intimate. Of course, the traditional costume is a must.

Castle Garden and Botanic Garden

The Castle was built between 1700 and 1704 by George William, Margrave of Brandenburg-Bayreuth and was the first baroque building in Franconia. Nowadays it is used by the university. The Castle Garden is a stunning place for a small walk and usually beautifully planted. It is one of the earliest baroque gardens established in Franconia. In Erlangen Botanical Garden, you can enjoy a beautiful variety of around 4000 species representing a wide range of plants of different climates. Established in 1626, it boasts a rich history and serves as a hub for research, education, and conservation efforts. From exotic tropical plants to native flora, the garden showcases a wide array of botanical treasures, making it a cherished destination for botanists, nature enthusiasts, and tourists alike.

Castle of Erlangen, Source: schulzfoto/stock.adobe.de

History

First named in 1002, Erlangen dates back a long time in history, a lot of which is visible in its beautiful old town and historic buildings. An event that still influences the city was the settlement of Huguenots after the withdrawal of the Edict of Nantes in 1685: The first french Huguenot refugees arrived in Erlangen in 1686.

Christian Ernst, Margrave of Brandenburg-Bayreuth, built a “new town” (Neustadt) for them. In 1706, the old town was almost completely destroyed by a fire, but soon rebuilt. In 1812, the old and new towns were merged into one. Nowadays life in Erlangen is mainly influenced by one of Germanys largest universities, the Friedrich-Alexander University Erlangen-Nuremberg and the Siemens technology group. Due to its large progress in medical research and industry, the metropolis of Erlangen-Nuremberg is called M*edical Valley* of Bavaria.

Franconian Switzerland, Source: schulzfoto/stock.adobe.de

Nuremberg, Bamberg and the Franconian Switzerland

Nuremberg, Bamberg, and the Franconian Switzerland region form a captivating trifecta in the heart of Bavaria, Germany. Nuremberg, a city steeped in history, is renowned for its medieval architecture, particularly the majestic Nuremberg Castle and the iconic Church of Our Lady. The city's rich cultural heritage is also evident in its bustling markets, such as the famous *Christkindlesmarkt*, and its significant role in the history of the *Holy Roman Empire*.

Bamberg, a UNESCO World Heritage Site, enchants visitors with its well-preserved medieval old town, adorned with charming half-timbered buildings and historic landmarks like the Bamberg Cathedral and the *Altenburg* Castle. Famous for its distinctive smoked beer, Bamberg offers a unique culinary experience amidst its picturesque cobblestone streets and scenic riverbanks.

Venturing into the *Franconian Switzerland* region, travelers are greeted by a landscape of rolling hills, lush forests, and dramatic rock formations. The name originated from romantic artists and poets from the 19th century that found resemblance in the landscape to Switzerland. Attractions like the *Devil's Cave (Teufelshöhle)* and the rock formations of the *Walberla* mountain *(Walberlafelsen)* add to the region's allure, making it a haven for outdoor adventurers and those seeking tranquility amidst nature's splendor. Together, Nuremberg, Bamberg, and the Franconian Switzerland region offer a captivating blend of history, culture, and natural beauty that continues to captivate travelers from near and far.

Key Note Speakers

As Key Note Speakers at ICVPB 2024, we are pleased to welcome:

Prof. Jennifer L. Long, MD, PhD

Dr. Jennifer Long is a surgeon-scientist in the Department of Head and Neck Surgery at the University of California, Los Angeles, where she also serves as the department's Vice Chair of Research. Her research focuses on regenerative medicine approaches to laryngeal disorders, and has been funded by the National Institutes of Health and the Department of Veterans Affairs. Dr. Long's clinical practice encompasses all aspects of laryngology and voice medicine, and she is a fellow of the American Laryngological Association. Before joining the faculty, she completed both residency in Head and Neck Surgery and fellowship in Laryngology at UCLA. Dr. Long earned her Medical Doctorate and a PhD in Chemical Engineering from the Medical Scientist Training Program at the University of Minnesota.

Prof. Jody Kreiman, PhD

Jody Kreiman is Professor in Residence of Head and Neck Surgery and Linguistics at the University of California, Los Angeles, where her research focuses on voice quality. Most recently she has proposed a psychoacoustic model of quality and has applied that model to examine the acoustic dimensions that meaningfully distinguish one talker from another. She received her PhD in Linguistics from the University of Chicago in 1987, and is a Fellow of the Acoustical Society of America and of the American Speech-Language-Hearing Association.

Prof. David A. Berry, PhD

David Berry has been a Professor of Head and Neck Surgery at UCLA for over 20 years. His research regarding basic mechanisms of regular and irregular vocal fold vibration has been funded by NIH/NIDCD for over 30 years. These mechanisms of vibration have been investigated using 3D high-speed imaging with various laboratory models of phonation, including silicone physical models of vocal fold vibration, the human excised larynx, and the in vivo canine larynx. Systematic studies of computer models of vocal fold vibration have also been employed to investigate these basic mechanisms. Among other things, his research has shown that many types of complex, irregular vocal fold vibrations, which appear high-dimensional in nature, can nevertheless be explained by just a few underlying modes of vibration, which are simply dis-entrained.

Prof. Dinesh K. Chhetri, MD

Dinesh Chhetri, is Professor of Head and Neck Surgery and Vice-chair for clinical operations at the University of California, Los Angeles (UCLA). He graduated from the UCLA School of Medicine, where he completed his residency in Otolaryngology and fellowship in Laryngology. He is an actively practicing Laryngologist and a voice scientist funded by the National Institutes of Health (USA). His main research interests are neuromuscular control of the larynx, evaluating the connections between laryngeal neuromuscular stimulation, vocal fold vibration, and voice quality. For nearly two decades his research has focused on laryngeal physiology and biomechanics using the in vivo canine model of phonation. After developing the technique for graded stimulation of individual intrinsic laryngeal muscles, the individual contributions of the muscles on phonatory posture, aerodynamics, acoustics, and vibratory behavior were investigated. Current investigations are focused on 3D reconstruction of glottal vibratory deformation, effects of laryngeal muscle activation on glottal volume waveform, as well as the trajectories of the medial surface landmarks in relation to neuromuscular stimulation and vocal fold microanatomy. He is serving as Council Member of the American Laryngological Association (ALA) and elected President of the ALA for 2023-2024.

Prof. Zhaoyan Zhang, PhD

Zhaoyan Zhang is a Professor of Head and Neck Surgery at the University of California, Los Angeles. His current research focuses on how changes in vocal fold physiology affect voice production and how to infer vocal fold physiology from the produced voice, leading toward clinical and speech technology applications. Dr. Zhang is a Fellow of the Acoustical Society of America, a recipient of the Quintana Award from the Voice Foundation, and an Associate Editor and Coordinating Editor for the Journal of the Acoustical Society of America.

Conference, Workshop and Social Event Locations

The conference will be hosted at the Novotel Erlangen Hotel:

Novotel Erlangen Hotel
Hofmannstrasse 34
91052 Erlangen
Deutschland

Workshop locations are listed below:

Workshop 1

Department of Materials Science and Engineering,
Institute of Biomaterials (WW7), Building II
Ulrich Schalk Straße 3
91056 Erlangen
Ground floor on the left

Workshops 2 and 3

Department for Artificial Intelligence in Biomedical Engineering,
Technical Faculty
Werner-von-Siemens-Straße 61
91052 Erlangen
Third floor to the right

Workshop 4

Institute of Polymer Technology (LKT)
Friedrich-Alexander-University Erlangen-Nürnberg
Am Weichselgarten 10
91058 Erlangen-Tennenlohe
Second Floor to the left

Workshop 5

University Hospital Erlangen
ENT Clinic/Department of Phoniatrics and Pediatric Audiology
Raumerstr. 1A
91054 Erlangen
CICERO Building - Seminar Room

Workshop 6 and 7

University Hospital Erlangen
ENT Clinic
Waldstr. 1
91054 Erlangen
ENT department - Lecture Hall and Laboratory

Workshop 8

University Hospital Erlangen
ENT Clinic/Department of Phoniatrics and Pediatric Audiology
Raumerstr. 1A
91054 Erlangen
CICERO Building – Conference Room

The Gala Dinner will be held at Bratwurst Röslein in Nuremberg:

Bratwurst Röslein
Rathausplatz 6
90402 Nürnberg

The Singing Evening will take place at Entlas-Keller Biergarten:

Entlas-Keller Biergarten GMBH
An den Kellern 5-7
91054 Erlangen

We will provide shuttle transportation for Workshops 1 and 4, departing from the conference hotel. The rest of the workshops and get togethers are in walking distance from the hotel, we will pick you up and walk you over. For the Gala Dinner, busses are waiting for you, that will depart from the Novotel Hotel Erlangen on Wednesday, July 26th at 17:30. For the singing evening, we will walk together, starting at the conference hotel.
We cordially invite you to join us. If you have any questions or concerns, please let us know.

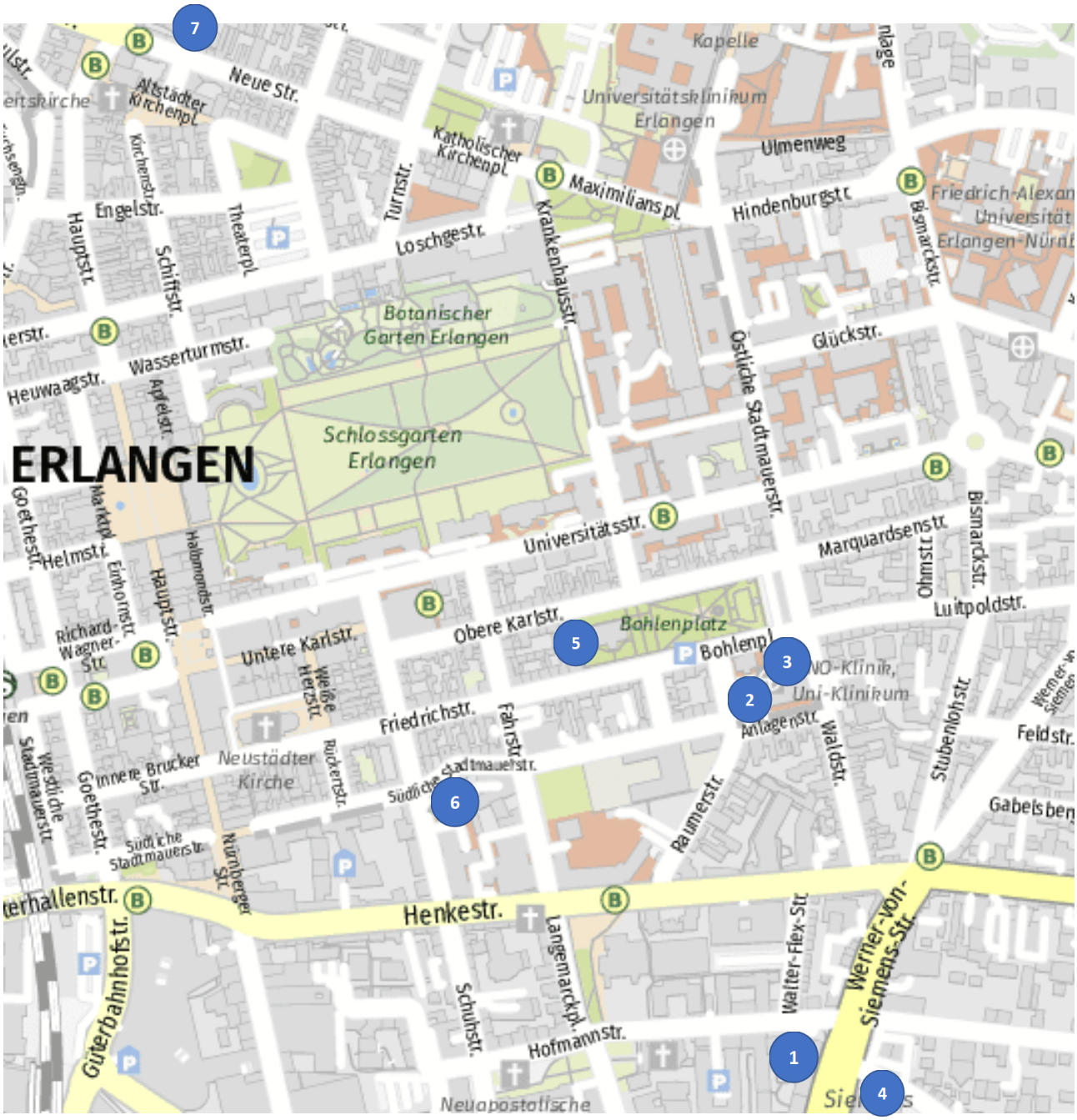
Get togethers – „Biergärten“:

Kitzmann Bräuschänke
Südliche Stadtmauerstr. 25
91054 Erlangen

Gaststätte Alter Simpl
Bohlenplatz 2
91054 Erlangen

Steinbachbräu Erlangen
Vierzigmannstr. 4
91054 Erlangen

Map of the Inner City



1: Novotel Hotel Erlangen (Conference Location)

2: Department of Phoniatrics and Pediatric Audiology (Workshops 5 and 8)

3: ENT-Clinic (Workshops 6 and 7)

4: Department for Artificial Intelligence in Biomedical Engineering

5: Restaurant "Alter Simpl"

6: Restaurant "Kitzmann Bräuschänke"

7: Restaurant "Steinbachbräu Erlangen"

Workshops

We will host eight workshops on state-of-the-art research concerning the wide field of laryngeal biomechanics. Workshops will be held before the main conference during July 22nd/23rd. The workshops will take place at the ENT department or at the respective institutes; transfer will be provided. The Locations are described on pages 7-8.

Monday, July 22nd

Workshops 1 and 2: morning

Engineering of Tailored Biomaterials for Tissue Regeneration:

- **Prof. Dr. Aldo Boccaccini**
  Dr. Rainer Detsch
- Location: Department of Materials Science and Engineering, Institute of Biomaterials (WW7), Building II, Technical Faculty, FAU
- From bioinks to living constructs, tissue regeneration through the use of bioactive materials, and an overview of biomaterials research in Erlangen

AI and Machine Learning Crash Course: Basics and Beyond

- **Prof. Dr. Katharina Breininger**
- Location: Department for Artificial Intelligence in Biomedical Engineering (AIBE), Technical Faculty, FAU
- Introduction to the fundamentals of AI, machine learning, and deep learning, highlighting their key distinctions, and best practices for your own research

Workshops 3 and 4: afternoon

Advanced Diagnostics by using AI methods

- **Prof. Dr. Andreas Kist**
- Location: Department for Artificial Intelligence in Biomedical Engineering (AIBE), Technical Faculty, FAU
- In-depth examination of AI solutions, within the voice and swallowing research domain; theoretical overview and hands-on session training your own deep neural network

Engineering and Analytics of Polymer Implants

- **Prof. Dr. Dietmar Drummer**
  MSc Samuel Schlicht
- Location: Institute of Polymer Technology (LKT), Technical Faculty, FAU
- Insights into the fundamentals of polymer processing and its impact on material and implant properties, Insights into methodologies for analyzing and testing chemical and physical characteristics of polymer-based implants

Tuesday, July 23rd

Workshops 5 and 6: morning

Females in Science: Forging your own Path as a Clinician/Researcher

- **Prof. Dr. Cate Madill**
- Location: ENT Clinic, Department of Phoniatrics and Pediatric Audiology, University Hospital Erlangen
- Identify your own unique strengths, preferences and attributes that will inform and support your career development in clinical science, research and/or academia

Experimental Larynx Models

- **PD Dr. Marion Semmler**
  PD Dr. Stefan Kniesburges
- Location: ENT Clinic, University Hospital Erlangen
- Learn about the characteristic advantages and disadvantages of the different types of models in theoretical and practical presentations, and explore the differences in the laryngeal anatomy of different species

Workshops 7 and 8: afternoon

Clinical Measurement of Professional Singers

- **Prof. Dr. Matthias Echternach**
  Dr. Marie Köberlein
- Location: ENT Clinic, University Hospital Erlangen
- Introduction to concepts and limitations of clinical voice measurement for professional singers

Numerical Larynx Models

- **Prof. Dr. techn. Stefan Schoder**
  Dr. Sebastian Falk
- ENT Clinic, Department of Phoniatrics and Pediatric Audiology, University Hospital Erlangen
- Insights on the complex challenges to model the phonation process numerically by methods of continuum mechanics, discuss concrete challenges and problems in simulating the process yourself

Complete Schedule

Sunday, July 21st		
18:00	**Optional First Get Together** - No registration needed	Kitzmann Bräu

Workshops

Monday, July 22nd		
Time	**Topic**	**Speaker/Organizer**
08:30 - 12:00	WS 1: Engineering of Tailored Biomaterials (Pick up at Novotel 08:00)	R. Detsch
09:00 - 12:00	WS 2: AI and Machine Learning: Basics (Pick up at Novotel 08:45)	K. Breininger
13:30 - 17:00	WS 3: Advanced Diagnostics Using AI (Pick up at Novotel 13:15)	A. Kist
13:30 - 17:00	WS 4: Polymer Implants (Pick up at Novotel 13:00)	D. Drummer
18:30	**Optional - Get Together** (Departure at Novotel at 18:20)	Alter Simpl

Tuesday, July 23rd		
Time	**Topic**	**Speaker/Organizer**
09:00 - 12:00	WS 5: Females in Science (Pick up at Novotel 08:45)	C. Madill
09:00 - 12:00	WS 6: Experimental Larynx Models (Pick up at Novotel 08:45)	M. Semmler
14:00 - 17:00	WS 7: Clinical Measurement of Singers (Pick up at Novotel 13:45)	M. Echternach
14:00 - 17:00	WS 8: Numerical Larynx Models (Pick up at Novotel 13:45)	S. Schoder
18:30	**Optional - Get Together** (Departure at Novotel at 18:10)	Steinbach Bräu

Main Conference

Wednesday, July 24th		
Time	**Topic**	**Speaker/Organizer**
08:00 - 08:08	Welcome	S. Kniesburges
08:08 - 08:15	Welcome from the Director of the University Hospital	H. Iro
08:15 - 09:00	**Key Note I**: Laryngeal and vocal tract control of voice production	Z. Zhang

09:00 - 10:00	**Artificial Intelligence I**	**Chair: T. Arias V**
From imaging to dynamics: a hybrid physics informed neural network *p.24*		X. Zheng
Forward mapping estimation for laryngeal motor control using machine learning techniques *p.26*		J. Parra
Evaluation of machine-learning pitch estimation algorithm *p.28*		T. Ikuma
Improved subglottal pressure estimation from neck-surface vibrations using transfer learning of deep neural networks trained from voice production model *p.30*		E. Ibarra, M. Zanartu

Coffee Break

10:30 - 12:15	**Artificial Intelligence II**	**Chair: A. Kist**
A deep learning based two-step approach for the detection of specific types of pig phonation *p.32*		P. Schlegel
Multi-Level Label Hierarchy for Disordered Voice Databases *p.34*		R. Gupta, C. Madill
Deep learning-based laryngotracheal and airway anatomical landmark detection for ultrasound-guided interventions *p.36*		G. Dion

Automatic detection and motion tracking of thyroid and arytenoid cartilages in dynamic translaryngeal ultrasound: a feasibility study for the diagnosis of vocal fold paralysis *p.38*	T. K. Bui
Do speech tasks influence machine learning performance in dysphonia detection? *p.40*	A. Yousef
Classification of vocal fold paralysis versus controls based on voice recordings with a Random Forest model: robustness validated using an external database *p.41*	J. Dindart
Objective assessment of functional dysphonia based on machine learning and high-speed videoendoscopy *p.43*	T. Schraut

Lunch Break

13:30 - 14:15	**Keynote II**: Vocal register transitions as a function of intrinsic laryngeal muscle stimulation	D. Berry
14:15 - 15:00	**Endoscopy**	**Chair: R. Samlan**
The use of 3D high-speed videoendoscopy for evaluating vocal fold vibratory characteristics as a function of frequency *p.44*		R. Patel
Reliability of segmentation of vocal fold edges from high-speed videoendoscopy and the parameters affecting it *p.46*		H. Ghasemzadeh
Comparing ten vocal fold closures using stroboscopy, high-speed videography, egg, and frequency analysis: a work in progress *p.47*		L. Popeil

Coffee Break

15:30 - 17:00	**Singing**	**Chair: N. Henrich**
Male and female voices show differences in spectral development over four years of conservatory training *p.48*		R. Walker
Articulatory and acoustic differences between lyric and dramatic singing in western classical music *p.49*		M. Echternach
Precision in articulators and phonation while performing staccato scales at different speeds compared to legato singing *p.50*		M. Köberlein
Examination of diaphragm control and its effects on pitch leap during opera singing using real-time MRI *p.52*		N. Toda
Acoustic contributions of the hypopharyngeal cavities in generating singer's formant: a case study *p.54*		H. Takemoto
Characteristics of glottal and supraglottal oscillations of ten different irregular phonation types during metal singing *p.56*		L. Traser

18:00	**Guided City Tour Nuremberg** (Departure Bus at Hotel to Nuremberg at 17:30)	
19:00 - 22:30	**Gala Dinner** (Departure Bus back to Erlangen at 22:45)	Bratwurst Röslein

Thursday, July 25th		
Time	**Topic**	**Speaker/Organizer**
08:00 - 08:45	**Key Note III**: Acoustic variability underlying pathological voice quality	J. Kreiman
08:45 - 10:00	**Disorders I**	**Chair: M. Gugatschka**
Detection of Common Voice Pathologies by Principal Component Analysis of Voice Acoustic Parameters *p.58*		C. Cantor Cutiva
Observation of laryngeal behaviors via high-speed videoendoscopy associated with the primary signs of ADLD *p.59*		K. Marks
Objective Quantification of Spasm Severity and Treatment Response in Adductor Laryngeal Dystonia with High-Resolution Manometry *p.61*		J. Hoffmeister
Relation between voice onset time, vocal hyperfunction type and voice quality: an exploratory study *p.63*		M. Brockmann-Bauser
Integrating biomechanical models, ambulatory monitoring, and statistical analysis for understanding non-phonotraumatic vocal hyperfunction *p.65*		C. Calvache

Coffee Break

10:30 - 12:15	**Disorders II**	**Chair: M. Brockmann-Bauser**
Adding vibrato-based ambulatory singing measures in the clinical assessment of patients with vocal hyperfunction *p.67*		E. Willis
Automated creak differentiates laryngeal dystonia and muscle tension dysphonia during a conversational speech task *p.69*		D. Dragicevic
Development of German continuous speech stimuli for relative fundamental frequency (RFF) *p.71*		M. Berardi
Influence of type I thyroplasty implant stiffness and location on the glottal pre-phonatory shape, stiffness and vibration *p.73*		W. Jiang
Can the vocal behavior of a mouse serve as a marker for vocal fold properties? *p.75*		T. Riede
Laryngeal and upper airway sequelae in post-acute influenza infection *p.76*		A. Foote
Exploring the effects of psychosocial stress on the laryngeal microbiome *p.77*		A. Venkatraman

Lunch Break

13:30 - 14:15	**Key Note IV**: Neuromuscular Control of Voice Production	D. Chhetri
14:15 - 15:30	**Disorders III**	**Chair: M. Kunduk**
Exploring the molecular basis of Reinke's edema *p.79*		M. Gugatschka
Differential cytokine response and altered DNA methylation in vocal fold fibroblasts of Reinke's edema patients *p.80*		M. Grill
Voice loudness, self-assessment, and auditory-sensory feedback in voice production in Parkinson's disease *p.82*		F. Contreras-Ruston
Early Alzheimer's disease-related laryngeal inflammation and muscle pathology in the TGF344-AD rat model *p.84*		M. Ciucci
Prodromal vocal changes in homozygous PINK1 knockout model of Parkinson disease *p.86*		M. Krasko

15:30 - 16:30	**Poster Session with Coffee**
~~#1:~~ cancelled	cancelled
#2: Investigating the Sound Generation in the Human Voice Based on Particle Image Velocimetry *p.153*	C. Näger
#3: Face Masks and Risk of Voice Disorder in University Professors *p.155*	A. Ghirardi
#4: Feature Selection of Phonation Parameters to Distinguish between Presbyphonia and Fes Treated Larynges from Young Larynges in an Ovine Model *p.157*	B. Jakubass

#5: A more Complete View into the Vocal Fold Tissue during Ex-vivo Phonation Experiments by Using a Combination of Pipette Aspiration and Ultrasound Elastography *p.159*	F. Scheible
#6: Optimizing Frequency Response in Ambulatory Voice Monitoring: Aligning Laboratory Accelerometer Data for Enhanced Voice Disorder Assessment *p.161*	J. Cortés
#7: Acoustic Characteristics of the Singing Voice and Vocal Tract Changes during Singing Instruction *p.163*	J. Takahashi
#8: Consideration of the Vocal Tract Acoustics of Different Voice Qualities with Regard to Voice Efficiency *p.165*	L. Traser
#9: Metagenomic Whole Genome Shotgun Analysis of the Airway Microbiome in Laryngotracheal Stenosis *p.167*	G. Dion
#10: Vocal Transient Analysis Using Piecewise Linear Approximation *p.169*	T. Ikuma
#11: The Sensitivity of Phonation Onset Pressure to Vocal Fold Shape *p.171*	S. Peterson
#12: Exploring Efficiency of Voice Acoustic Parameters for Detecting Voice Disorders: a Systematic Review of Literature *p.173*	A. Yousef
#13: Influence of Supraglottal Tract on Flow Separation Vortices during Vocal Fold Closing in Human Phonation *p.174*	W. Jiang
#14: Real-time Voice Quality Alteration: Tackling the Complexity of Simulating Dysphonia *p.176*	I. Schiller
#15: Notes on Cascading Simplifications of the Vocal Tract and its Relevance for the Transfer of Acoustic Energy *p.177*	M. Fleischer
#16: Acoustic and Neurophysiological Aspects of Lombard Effect *p.178*	L. Rivera
#17: Spatial Transcriptomics of the Vocal Fold Tissue of Intact and Ovariectomized Female Rats *p.180*	P. Sivasankar
#18: The Influence of Cigarette Smoke-stimulated Vocal Fold Fibroblasts on Inflammatory Cytokine Release of Macrophages *p.182*	B. Steffan
#19: Amniotic Fluid as a Potential Treatment Following Vocal Fold Injury *p.184*	B. Christensen
#20: Extralaryngeal Surface EMG Classification during Sentence Production for Vocal Fatigue Detection Using GA-SVM for Confounder Removal *p.185*	Y. Gao
#21: Does Forced Whisper have an Impact on Voice Parameters? *P.187*	T. Pilsl
#22: The Prevalence of Creak Across Breath Groups in Speakers with and Without Adductor Laryngeal Dystonia *p.188*	K. Marks

16:30 - 18:30	**Physiology**	**Chair: R. Patel**
Ideal glottal waveform: basic concepts and kinematic modelling with synthetic kymograms *p.88*		J. G. Svec
How do multiple sound sources in the airway interact? *p.90*		I. Titze
Source Filter Synchronization in Connected Speech Samples *p.91*		L. Maxfield
Influence of subglottal vibrational excitation in the range of 2-6 kHz on voice parameters *p.93*		P. Hoyer
Intralaryngeal neural networks, a novel avenue in neurolaryngology and its impact on phonatory posturing *p.95*		R. Tracicaru
Exploring nonlinear phenomena (NLP) in animal vocalisations through oscillator theory *p.97*		H. Herzel
Does sex matter? In vitro exposure of vocal fold fibroblasts to vibration *p.98*		T. Grossmann
Metabolic validation of an anaerobic vocal demand task: novel application of blood lactate as a biomarker of vocal function *p.99*		M. Morton-Jones

19:00	**Singing Evening** (Departure at Novotel 18:40)	Entlas Keller

Friday, July 26th		
Time	**Topic**	**Speaker/Organizer**
08:00 - 08:45	**Key Note V**: Cell-based outer vocal fold replacement for vibratory tissue repair	J. Long
08:45 - 10:00	**Bio-Tissue Engineering**	**Chair: N. Li-Jessen**
Single-cell atlas of vocal fold and laryngeal embryogenesis, maturation and aging *p.100*		T. Lunga
Minimally invasive in situ bioprinting for vocal fold regeneration using a cable-driven continuum manipulator *p.102*		S. Groen
The potential role of SP-G and PLUNC in wound healing and tumor pathogenesis in the human larynx respectively the vocal fold *p.104*		A. Scheer
LRIG1 cells as epithelial stem cells for vocal fold maintenance and regeneration *p.105*		V. Lungova
Parameter optimization of agent-based models for vocal fold biomaterial design *p.107*		N. Li-Jessen
Coffee Break		
10:30 – 11:45	**Computational Modeling**	**Chair: C. Fabrouz**
Perception of voicing in speech produced by adult and pediatric kinematic vocal fold models *p.109*		E. Heller Murray
Modelling the Influence of Localized Edema on Vocal Fold Kinetics and Kinematics *p.111*		S. Peterson
Effects of false and aryepiglottic fold constrictions on voice production in a simplified vocal tract *p.113*		T. Yoshinaga
Computational analysis of vocal fold posturing in asymmetric laryngeal muscle activations *p.115*		Q. Xue
Vocal3D: fully automatic 3D reconstruction of phonating human vocal folds via structured light laryngoscopy *p.117*		J. Heningson
Lunch Break		
13:00 - 15:00	**Experimental Modeling**	**Chair: L. Oren**
FSAI-01: A benchmark dataset for aeroacoustic simulations of human phonation *p.119*		S. Schoder
Experimental modelling of the influence of vocal folds compliance on human vocal tract acoustic properties *p.121*		V. Radolf
Vocal Fold Reconstruction from Optical Velocity and Displacement Measurements *p.123*		C. Näger
Experimental measurement of the internal tissue velocity in a self-oscillating synthetic vocal fold model *p.125*		B. Erath
Intraglottal pressure and contact pressure distribution in excised human larynges *p.127*		S. Lehoux
Vibratory response of an adaptive synthetic larynx model *p.129*		B. Tur
Highly dynamic rotational molding of thin-walled larynx models for fluid dynamic modelling *p.131*		S. Schlicht
Flow-induced oscillations of a vocal-fold replica with tailored anterior-posterior and medial-lateral articulations *p.133*		M. Tlaidi
Coffee Break		
15:30 - 17:45	**Therapy and Diagnostics**	**Chair: M. Echternach**
Voice onset and offset oscillatory patterns based on place, manner and voiced/voiceless consonant environment *p.135*		M. Kunduk
Evaluation of intrinsic laryngeal muscle activity using surface electromyography *p.137*		X. Yang
Investigating the correlation between diverse vocal tasks and intermuscular coherence: a pilot study using high-density electromyography *p.139*		J. Martinez
Dynamic 3D Magnetic Resonance Imaging of the Vocal Fold Oscillation *p.141*		J. Fischer
Evaluation of the frequency-dependent young's modulus of the vocal folds by optical coherence tomography in combination with pipette aspiration *p.143*		R. Lamprecht

Vocdoc – a novel app to capture long-term voice variations in the wild *p.145*		F. Pokorny
Effect of thyroplasty type I implant location on glottal medial shape during phonation of a canine larynx model *p.147*		C. Farbos de Luzan
The inflammatory and biomechanical effects of inhaled corticosteroids on rabbit vocal fold tissue *p.149*		B. Christensen
Investigating the influence of the vocal tract acousto-mechanical resonance on the dynamics of semi-occluded vocal tract exercises involving tubes immersed in water *p.150*		A. Ricardo da Silva
17:45 - 18:00	**Goodbye**	S. Kniesburges, M. Döllinger

Gala Dinner and Singing Evening

Gala Dinner on Wednesday 26th

The Gala Dinner will be hosted at *Bratwurst Röslein*, the biggest Bratwurst Restaurant in the world! Founded in 1431, *Bratwurst Röslein* comes from a long tradition of authentic Franconian cuisine and invites you to enjoy an evening with local specialties in its rustic interiors. Located in the heart of the old town in Nuremberg, you will also have the opportunity to explore the historical city of Nuremberg prior to the dinner. During a one-hour guided tour, you will gain interesting insights into the history of this diverse city while simultaneously enjoying beautiful views of the old buildings and city skyline.

We will offer bus transportation to and back from the event, so you can simply enjoy the dinner. We are very much looking forward to an unforgettable evening

17:30 - Bus Departure at Novotel
18:00 - Guided City Tour through Nuremberg
19:00 - Gala Dinner
22:45 - Bus Departure back to Novotel

Singing Evening on Thursday 27th

We are also offering a singing and music event, kindly organized by Nathalie Henrich Bernardoni. The Singing Evening will take place at *Entlas Keller Biergarten* in Erlangen. This beautiful location has a stunning patio and has very old beer cellars in Erlangen's *Burgberg*, which date back to 1686 and are still in use. You can enjoy the unique beer garden atmosphere outside and sample homemade food and beers. The highlight of the evening is our english guided tour through the historic cellars and the one-hour singing event in one of the old ice chambers deep inside the mountain, which offers fantastic acoustic conditions. These ice chambers were once used to cool the beer during spring and summer before refrigerators existed. Of course, we can also make music outside!

18:40 - Departure at Novotel
19:00 - Singing Evening

General information

Procedure in general:

Talks: Presentations are scheduled for **12 minutes** with **3 additional minutes for questions**.

Poster: The poster session is scheduled **for Thursday 25th between 15:30 and 16:30.** We ask all authors to put up their posters on Thursday morning **until 09:00 am**.

Key Note Speakers: The plenary talks of the keynote-speakers are scheduled **30 minutes** with **15 minutes for questions**

Lunch: Lunch as well as coffee for the coffee breaks will be served at the conference venue for you and is included in the registration fees.

Presenter Instructions:

To allow a trouble-free rotation, **please hand us your presentations between 7:00 – 7:45** am or during coffee/lunch breaks at day of presentation or the day before. Make sure to convert it into **PDF or PowerPoint-file**! No private laptops can be connected. Thanks a lot for your cooperation.

We thank all participants for their contributions!

All abstracts can be found in the same order as illustrated in the conference schedule.
If you have any questions, we are always happy to help!

Abstracts

TALKS

FROM IMAGING TO DYNAMICS: A HYBRID PHYSICS INFORMED NEURAL NETWORK MODEL FOR SUBJECT SPECIFIC PHONATION SIMULATION

Biao Geng[1], Qian Xue[1], Xudong Zheng[1], Michael Dölinger[2]

[1] Department of Mechanical Engineering, Rochester Institute of Technology, Rochester, New York, USA

[2]Division of Phoniatrics and Pediatric Audiology at the Department of Otorhinolaryngology Head & Neck Surgery, FAU Erlangen-Nürnberg, Germany

***Keywords*: Physics-informed neural network; Inverse problem; Modal dynamics;**

Abstract:

Introduction:

The three-dimensional (3D) tissue dynamics of the vocal folds play a critical role in phonation and could provide information on pathological conditions. However, it is extremely difficult to measure the 3D dynamics during clinical examination. On the other hand, two-dimensional (2D) endoscopic imaging of the vocal fold vibration can be routinely obtained, and the 2D glottal shape can be procedurally segmented (see e.g. Ref. [1]). In this study, we aim to inversely determine the 3D tissue dynamics from 2D vocal fold edge segmentation. We designed a novel hybrid physics informed neural network (PINN) based differentiable learning algorithm that integrates a recurrent neural network (RNN) model of 3D continuum soft tissue with a differentiable fluid solver. The learning algorithm can infer the 3D flow-induced vocal dynamics and other physical quantities from high speed video endoscopy. Drawing upon prior knowledge of tissue material properties, our algorithm has demonstrated success in inferring 3D vocal dynamics on subject-specific voice production problems [2]. In the present study, we further introduce a differentiable eigen-solver into the learning algorithm to inversely determine the material properties of vocal fold tissues, which eliminates the need for precomputing eigenmodes and extends its applicability to *in-vivo* experimental and clinical studies. This study contributes to our overarching goal of developing a new AI-enabled data-assimilation computational framework that enables integration of multimodal experimental/clinical data and high-fidelity physics-based subject-specific modeling to provide accurate, realistic, robust, efficient and reliable simulation-based evaluation of individual vocal system.

Methods:

Figure 1 illustrates the overall structure of the simulation framework and the methodology. The central idea of PINNs is to use physics to inform the network training by penalizing the violation of physical laws and constraints, thus enabling sparse-label learning, assimilation of indirect data, and improved sample efficiency. Here, the VF is modeled using linear finite element method and its vibration is solved in the truncated eigen-space. The RNN model takes the sequence of segmented glottal shapes as the only input and infers the modal coefficients of the eigenmodes. The RNN model also infers the material property, which is used to compute the eigenmodes with a differentiable eigenmode solver. Then, the full predicted 3D VF deformation is constructed from the eigenmodes and the modal

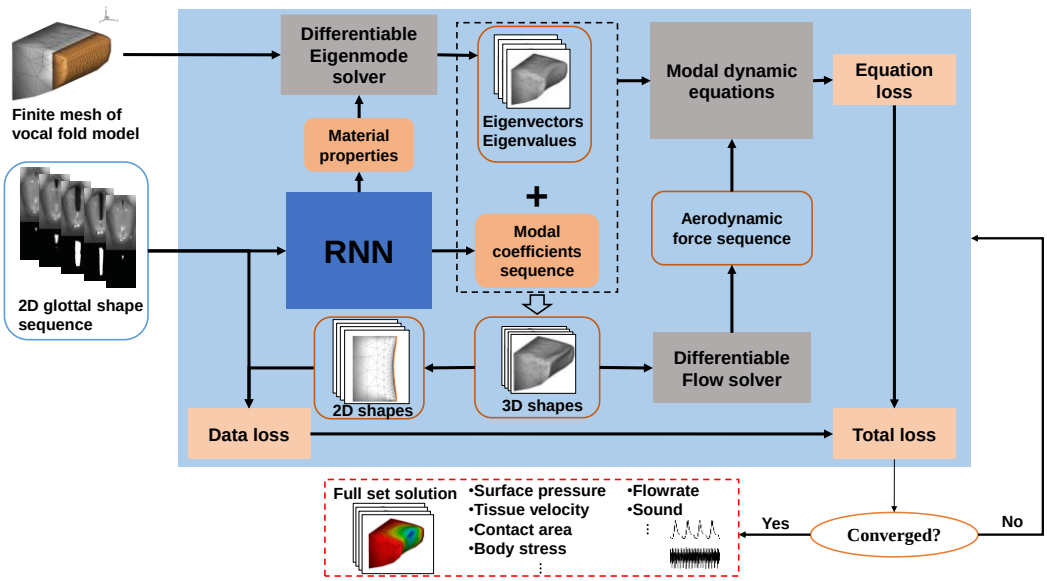


Figure 1. Structure of the discrete PINN-differentiable programming algorithm. The illustration uses a simplified VF model.

coefficients. With this, the residual of the equations of motion is calculated with a reduced-order flow solver and is incorporated into the loss function of the network as equation loss. Additionally, the 2D glottal shape procedurally generated from the predicted 3D deformation is compared against the input and incorporated into the loss function as data loss. The learning algorithm uses the auto differentiation framework of PyTorch [3] to minimize the loss function.

Results:

The algorithm was tested on a synthetic dataset using a simplified 2-layered VF model. The vocal fold model was extruded from a profile (Figure 2(a)). Figure 2(a)-(c) show the process of synthetic input generation. The 3D vocal fold vibration was generated from flow-structure interaction simulation. 2D glottal shapes were generated from the deformed vocal fold from top-view projection. For the network input, 16 glottal shape samples per vibration cycle over two cycles were used. Figure 2(d) shows the convergence of total loss for a typical case with eigenmode solution. Figure 2(e) demonstrates the algorithm's capability to inversely determine the material properties of the vocal fold. In these case, the only inferred material parameter is the cover layer modulus. The network inference converges to the true value (blue line) regardless of the initialization (diamond) that spans two orders of magnitude. It was observed that the solution was insensitive to material properties of the body layer which had much less significant effect on the glottal shape than the cover layer. Testing on a broader range of material properties and more complex material models will be performed. Figure 2(C) shows the prediction error of the 3D vocal fold shapes over one vibration cycle. The error is slightly larger than the previous results without eigenmode solution[2].

Conclusions:

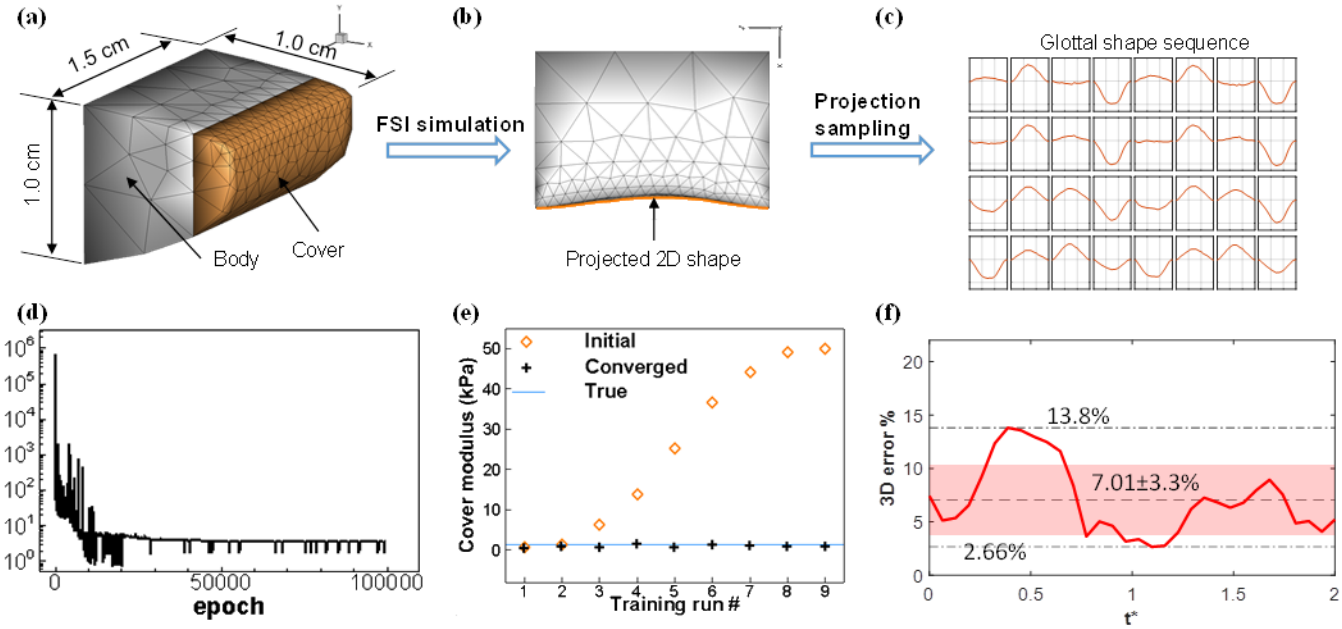


Figure 2. (a) VF model. (b) Top-view projection. (c) Glottal shape sequence for network input. (d) Convergence of network in 100,000 epochs. (e) Convergence of cover modulus. (f) Error of 3D prediction over one cycle.

A hybrid PINN-based learning algorithm is being developed to infer 3D VF tissue dynamics from only 2D endoscopic imaging without knowing the tissue material properties *a priori*. Results show that the algorithm can successfully reconstruct the 3D motion of vocal fold and estimate other features such as flow rate and acoustic signals, which are difficult to measure *in vivo*. Current application is constrained by the computational speed. The algorithm also assumes that the mechanical status of postured VFs can be represented by an equivalent stiffness of soft tissue, which may need revision for further clinical application.

References:

[1]. Kist et al. Rethinking Glottal Midline Detection. Sci Rep 2020, 10, 20723.
[2] Movahhedi et al. Predicting 3D Soft Tissue Dynamics from 2D Imaging Using Physics Informed Neural Networks. Communications Biology 2023, 6, 541.
[3] Paszke et al. Pytorch: An Imperative Style, High-Performance Deep Learning Library. Advances in neural information processing systems 2019, 32.

FORWARD MAPPING ESTIMATION FOR LARYNGEAL MOTOR CONTROL USING MACHINE LEARNING TECHNIQUES

Clara Sorolla[1], Jesús A. Parra[2,1], Emiro Ibarra[2,1], Gabriel Alzamendi[3,4], Matías Zañartu[2,1]

[1] Advanced Center for Electrical and Electronic Engineering, Universidad Técnica Federico Santa Maria, Valparaíso, Chile

[2] Department of Electronic Engineering, Universidad Técnica Federico Santa Maria, 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*: Speech-Motor Control; Machine Learning; Computational Modeling**

Abstract:

Objectives / Introduction:

Understanding the factors that influence vocal function is a complex task. A recent major development in this field is the introduction of LaDIVA [1], a neurocomputational model that facilitates laryngeal motor control in speech acquisition and production. This study aims to refine the forward mapping process within LaDIVA by incorporating a more sophisticated voice production model, coupled with an efficient machine learning-based implementation. These enhancements are vital for accurately computing the outputs for the auditory feedback controller, thereby not only improving the computational efficiency of the LaDIVA model but also illustrating the potential of machine learning in developing individualized representations of laryngeal motor control.

Methods:

The function we aim to replace enables a direct mapping between various vocal function features. The regression model takes as its inputs the activation of the cricothyroid muscle (CT), the activation of the thyroarytenoid muscle (TA), and the subglottal pressure (P_S). Its outputs are fundamental frequency (f_o) and sound pressure level (SPL). Initially, this function was a lookup operation within a pre-generated table of features using the Body-Cover Model (BCM) [2]. After evaluating various prevalent machine learning models, the multiple output Random Forest Regressor (RFR) demonstrated an aptitude for effectively handling this specific type of data.

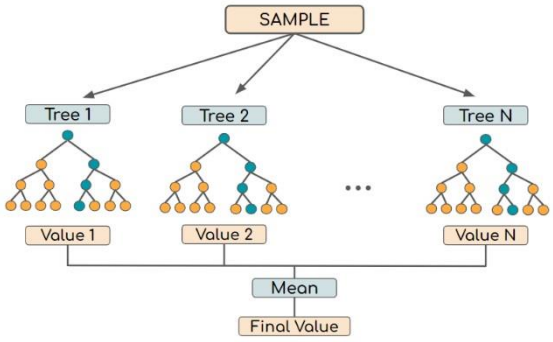


Figure 1: Random Forest Regressor Architecture.

The Random Forest Regressor (RFR), depicted in Figure 1, operates as a bagging ensemble model that integrates multiple decision trees. A decision tree is a hierarchical structure composed of decision nodes, which facilitate branching based on binary (true or false) conditions, culminating in terminal nodes that produce the final result. This structure allows for the progressive subdivision of data into increasingly homogeneous subsets. The RFR begins by randomly distributing the data into subsets, each of which is used to train an individual decision tree. Then, each tree calculates its own output, and the individual outputs are averaged at the end. This approach effectively mitigates the overfitting typically associated with a single decision tree by incorporating the outputs of many trees, thereby enhancing the model's reliability and accuracy.

Using this technique instead of the previous LaDIVA lookup table is less time-consuming and adds more flexibility to the function, as it does not require all values to be predefined to correctly perform the regression. This approach also facilitates retraining the model to adapt to a different subject with another set of characteristics.

To demonstrate the adaptability of this technique to individual subjects, various simulation sets were created, with each set differing based on the sex of the subject, being either female or male, and on the resting length of the vocal folds. To generate the simulations the triangular body-cover model (TBCM) was used [3]. The TBCM is a six mass-spring vocal fold model that

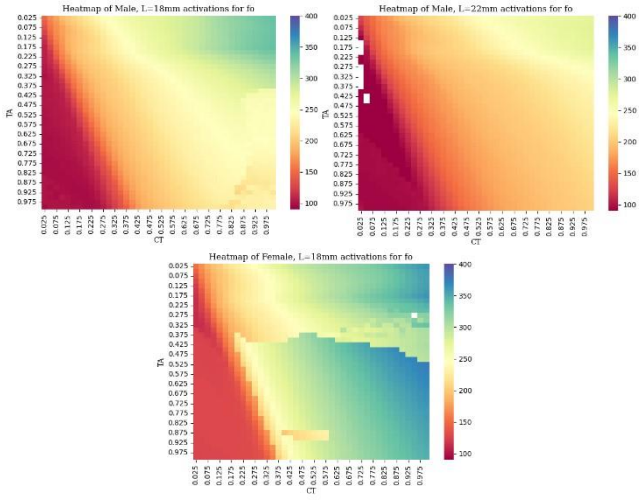


Figure 2: Heatmaps for fo of the different data sets with a lung pressure of 1500 Pa.

simulates phonation using a triangular glottis shape and five intrinsic laryngeal muscles for a more realistic vocal fold posturing and laryngeal motor control.

Results:

Three datasets comprising approximately 30k samples each were generated through multiple simulations using the TBCM. One dataset, designed to represent a typical female, featured a vocal fold length of 18 mm, while the other two datasets, intended to mimic typical males, had vocal fold lengths of 18 mm and 22 mm, respectively.

Figure 2 displays a heatmap of f_o, highlighting the distinctions among these datasets. Each dataset was utilized to train four distinct Random Forest Regressors, with 80% of the data allocated for training and the remaining 20% for validation purposes. Table 1 provides a summary of the performance metrics for each model. Figure 3 illustrates the outcomes from one of these regressors, showcasing two scatter plots that compare the outputs estimated by the RFR with the simulated references and their corresponding empirical error distributions. Note that the performance of the RFR was considerably more efficient than the previous function used in LaDIVA. For instance, when evaluating 300 samples the computation time was reduced from 35ms to less than 5ms, demonstrating the high computational efficiency of this technique in contrast to using a lookup table.

Table 1: Results of the three Random Forest regressors.

Dataset	SPL		fo	
	MAE	**R^2**	**MAE**	**R^2**
Female (L=18 mm)	0.5680	0.9753	2.3411	0.9901
Male (L=18 mm)	0.1965	0.9954	0.8391	0.9992
Male (L=22 mm)	0.1276	0.9950	0.4869	0.9997

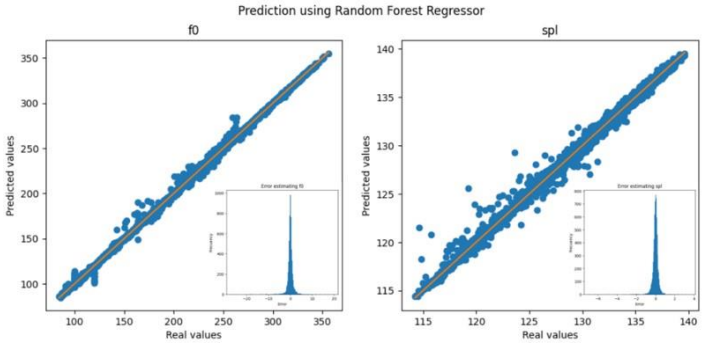


Figure 3: Results of Random Forest Regressor trained with the simulations of a male with a vocal fold length of 18 mm.

Conclusions:

After observing the results, we can conclude that the RFR is an adequate tool for the forward mapping estimation of speech-motor control in the LaDIVA model. Its ability to correctly adapt to different datasets will significantly contribute to building subject-specific representations. This opens a new path to using other techniques that are compatible with Machine Learning, such as Transfer Learning, in the case of scarce data when using *in vivo* measurements.

Acknowledgements:

This research was supported by the National Institutes of Health (NIH) National Institute on Deafness and Other Communication Disorders grant P50 DC015446, and ANID programs FONDECYT 1230828, BASAL FB0008, and Becas de Doctorado Nacional 21190074 and 21202490. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References:

[1] H. R. Weerathunge, G. A. Alzamendi, G. J. Cler, F. H. Guenther, C. E. Stepp, and M. Zañartu, "*LaDIVA: A neurocomputational model providing laryngeal motor control for speech acquisition and production*," *PLoS Comput Biol*, vol. 18, no. 6, Jun. 2022.

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

G. A. Alzamendi, S. D. Peterson, B. D. Erath, R. E. Hillman, and M. Zañartu, "*Triangular body-cover model of the vocal folds with coordinated activation of the five intrinsic laryngeal muscles*," *J Acoust Soc Am*, vol. 151, no. 1, pp. 17–30, Jan. 2022.

EVALUATION OF MACHINE-LEARNING PITCH ESTIMATION ALGORITHMS

Takeshi Ikuma, PhD[1,2], Andrew J. McWhorter, MD[1,2], Melda Kunduk, PhD[1,2,3]

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

Keywords: Disordered Voice; Acoustic analysis; Pitch; Subharmonics

Abstract

Objectives / Introduction:

This presentation aims to demonstrate the performance of existing machine-learning (ML) based pitch detectors, especially for handling so-called subharmonic errors. Accurate pitch estimation is vital for objective voice analysis. Most acoustic parameters rely on either periods or fundamental frequencies f_o of vocal cycles with exception of only a handful parameters (e.g., cepstrum peak prominence and loudness). In a recent study[1], ML algorithms—CREPE[2] and FCN-F0[3]—were shown to outperform the Praat pitch analysis algorithm[4], especially in estimating f_o of the voices of head and neck cancer patients. However, the study did not reveal what type of errors these ML algorithms mitigate. The present study was focused on analyzing a particular type of pitch estimation error: incorrect selection of an integer-divisor of the true f_o as the estimated. We call this type of error a subharmonic error because it is registered mostly when a voice signal is nonmodal and contains subharmonics. The challenge here is that the true fundamental frequency of a subharmonic signal is indeed an integer-divisor of the speaking f_o, i.e., the pitch intended by the speaker. As such, a pitch detector cannot solely depend on the periodicity of the signal and requires additional information to make a correct choice.

Methods:

***Data*:** All 709 sustained /a/ recordings of KayPENTAX Disordered Voice Database [5] were included. While each recording was processed at once, pitch estimates in a 50-millisecond (ms) interval were pooled for the analysis, yielding 16174 total sample points (frames).

f_o **Annotation.** The truth values for the fundamental frequency f_o in all signal intervals were evaluated in three steps. First, the initial estimates were gathered from the Praat. Then, these estimated f_o's were reviewed, and those which Praat incorrectly estimated were adjusted manually with a custom computer program. Finally, the estimates were refined using the time-varying harmonic model with a gradient-based optimization[6].

Algorithms. *Praat* pitch detector is based on the autocorrelation function and uses the hidden Markov model (HMM, an unsupervised machine-learning technique) as the postprocessor to correct the errors. The default configuration was used except for the minimum pitch was adjusted to 60 Hz to handle the lowest f_o present in the dataset. This results in Praat to set its analysis frame size to be 50 ms. The recordings were resampled to 8 kHz first. The *CREPE* algorithm[2] uses a six-layer convolutional neural network (CNN) model. The model signals to be resampled to 16 kHz with input signal frame size of 1024 samples (64 ms; 28% frame overlap). The original model coefficients were used. The *FCN-F0* algorithm FCN-F0 algorithm[3] is an extension of the CREPE with a 7-layer model, taking input signals sampled at 8-kHz. Specifically, the FCN-933 model with input frame size of 993 (124 ms; 148% frame overlap) and the original model coefficients were used. Finally, the best Praat pitch candidates before postprocessing (i.e., the estimates of an autocorrelation-function based algorithm, *ACF*) was evaluated as a reference without employing any machine learning techniques.

Performance Metric. To detect subharmonic errors accurately, the output of each algorithm was refined using the time-varying harmonic model as the annotated truth. The refined estimate was recorded as f_o. In other words, the algorithm output was considered correct if it yields the same harmonic model as the annotated. The ratio of the truth and estimated, f_o/f_o, was used as the performance metric.

If $f_o/f_o \approx 1$, the estimated f_o matches the truth (labeled "Correct"). On the other hand, if f_o/f_o is approximately an integer greater than 1, the algorithm committed a subharmonic error (labeled "Subh"). Note that this error does not guarantee that the signal contains subharmonics as the error can also be caused by chance. Finally, non-integral f_o/f_o values indicate either that the estimator picked an f_o with no apparent relationship to f_o or that the truth f_o was too aggressively annotated (labeled "Other"). To account for a numerical error, f_o/f_o value within ±0.01 of an integer was treated as an integral outcome, i.e., either "Correct" or "Subh". **Results:**

Out of 16174 frames, 15956 (98.7%) were found to contain periodic behavior with annotated f_o's. Based on the f_o/f_o metric, Table 1 shows the overall performance of the pitch detectors. It is apparent that the ACF is prone to making subharmonic errors (34.5%) and Praat's HMM successfully eliminated 75.2% of these errors. For these two algorithms, the subharmonic errors are dominant over the other types of errors. Meanwhile, both CNN solutions clearly outperform the former two with above 95% correctness. More importantly, they demonstrated their resiliency to the subharmonic errors: reduced the amount of subharmonic error by 94.3% for CREPE and 96.1% for FCN-F0

relative to ACF. The CREPE and FCN-F0 committed fewer subharmonic errors than the other types. A minor downside of both CREPE and FCN-F0 was that they introduced a few frequency-doubling errors ($f_o/f_o = 0.5$; 17 for CREPE and 11 for FCN-F0) which Praat had none of and ACF had in 3 frames.

Table 2 shows how Praat, CREPE, and FCN-F0 improved over ACF. The top row of the table reveals that the CNN algorithms selfdom made errors on the frames which the ACF was correct (<1.0%), especially the subharmonic errors (<0.2%). In comparison, the HMM postprocessor of Praat incorrectly flipped 1.5% of the correct ACF estimates to subharmonic errors.

Table 1: Pitch Detection Outcomes (occurrences)

	ACF	Praat	CREPE	FCN-F0
Correct	9830 (61.6%)	13931 (87.3%)	15200 (95.3%)	15335 (96.1%)
Subh	5506 (34.5%)	1368 (8.6%)	316 (2.0%)	217 (1.4%)
Other	620 (3.9%)	657 (4.1%)	440 (2.8%)	404 (2.5%)
Subh, % rel ACF	100.0%	24.8%	5.7%	3.9%

Table 2: Contingency matrix between ACF and other algorithms (occurrences)

ACF	Praat			CREPE			FCN-F0		
	Correct	Subh	Other	Correct	Subh	Other	Correct	Subh	Other
Correct	9646	152	32	9740	15	75	9762	3	65
Subh	4269	1209	28	5239	247	20	5315	179	12
Other	16	7	597	221	54	345	258	35	327

Conclusions:

The CNN-based pitch detectors—FCN-F0 and CREPE—have demonstrated dominant performance over the pitch detector of the widely used Praat software by correctly detecting the pitches on over 95% of the sustained /a/ audio frames. Remarkably, these detectors are driven by model coefficients which were trained with synthesized data (CREPE) or with speech database without any voice disorder samples (FCN-F0). The CNN detectors especially excelled in avoiding committing subharmonic errors likely by learning the details of the cyclic behavior of the signals beyond their periodicity. These CNN (and other deep learning) models could improve their performance further by training them with clinically relevant data and by using a postprocessing technique similar to Praat.

References

1. Vaysse R, Astésano C, Farinas J. Performance analysis of various fundamental frequency estimation algorithms in the context of pathological speech. *J Acoust Soc Am*. 2022;152(5):3091-3101. doi:10.1121/10.0015143
2. Kim JW, Salamon J, Li P, Bello JP. Crepe: A Convolutional Representation for Pitch Estimation. In: *2018 IEEE Int. Conf. Acoust. Speech Signal Process. ICASSP*.; 2018:161-165. doi:10.1109/ICASSP.2018.8461329
3. Ardaillon L, Roebel A. Fully-Convolutional Network for Pitch Estimation of Speech Signals. In: *Interspeech 2019*.; 2019:2005-2009. doi:10.21437/Interspeech.2019-2815
4. Boersma P. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Proc Inst Phonet Sci*. 1993;17:97-110.
5. KayPENTAX, Massachusetts Eye and Ear Infirmary. Disordered Voice Database and Program [Model 4337]. Published online 2006.
6. Ikuma T, Story B, McWhorter AJ, Adkins L, Kunduk M. Harmonics-to-noise ratio estimation with deterministically time-varying harmonic model for pathological voice signals. *J Acoust Soc Am*. 2022;152(3):1783-1794. doi:10.1121/10.0014177

IMPROVED SUBGLOTTAL PRESSURE ESTIMATION FROM NECK-SURFACE VIBRATIONS USING TRANSFER LEARNING OF DEEP NEURAL NETWORKS TRAINED FROM VOICE PRODUCTION MODEL

Emiro Ibarra[1], Julián Arias[2], Juan Godino[2], Daryush Mehta[3], Matías Zañartu[1]

[1] Department of Electronic Engineering, Universidad Técnica Federico Santa María, Valparaíso, Chile
[2] Bioengineering and Optoelectronics lab (ByO), Universidad Politécnica de Madrid, Madrid, Spain
[3] Center for Laryngeal Surgery and Voice Rehabilitation Laboratory, Massachusetts General Hospital, MA, USA

***Keywords*: Neural Network; Subglottal Pressure Estimation; Transfer learning; Voice Production Model**

Abstract:

Introduction:

The ambulatory assessment of subglottal air pressure (Ps) is vital for advancing the clinical management of voice disorders. This parameter regulates vocal fold oscillation dynamics, modulates loudness, and adjusts the fundamental frequency. It is also closely linked to phonatory efficiency and vocal effort. However, the direct measurement of Ps typically requires invasive or specialized techniques, such as tracheal puncture or circumferentially vented mask, which limits its practical application in clinical and ambulatory settings.

In our previous work [1], we proposed a non-invasive method to estimate Ps, among other parameters, using necksurface vibrations recorded by an accelerometer (ACC). The approach employed a Neural Network (NN) trained on data simulated from physiological numerical models, achieving high performance with these synthetic datasets. However, we observed a significant decrease in performance when this method was applied to predict Ps using an accelerometer sensor. In our current research, we aim to improve this by applying transfer learning (TL) to a neural network regression initially trained on voice production model data. This strategy involves retraining the NN using laboratory datasets and sequentially freezing the hidden layers of the neural network, thereby enhancing the accuracy of subglottal pressure estimations derived from ACC data.

Methods:

In vivo measurements: The data consisted of synchronous laboratory recordings from intraoral pressure (IOP), oral airflow volume velocity (OVV), a microphone (MIC), and an ACC from 79 adult female subjects who were vocally healthy. The participants produced strings of /pæ/ syllables at three levels of loudness: comfortable, loud, and soft. Reference values for Ps were obtained from the IOP measurements. The ACC-based glottal airflow was estimated using the impedance-based inverse filtering (IBIF) algorithm. Then, the middle 50 ms of the glottal airflow signal, estimated from IBIF, was selected to compute six acceleration-based aerodynamic features: unsteady glottal airflow (ACFL), maximum flow declination rate (MFDR), open quotient (OQ), speed quotient (SQ), spectral tilt (H1−H2), and fundamental frequency (f_0). The sound pressure level (SPL) was obtained from the MIC signal.

Synthetic voice dataset: The synthetic dataset comprises 13,000 sustained phonation simulations, generated using a muscle-controlled voice synthesizer. This synthesizer incorporates a triangular body-cover vocal fold model and enables a three-way interaction at the glottal level between sound, flow, and vocal fold tissue. The dataset encompasses a wide range of variations in model control parameters, such as lung pressure and muscle activation levels [1]. Each simulation had a duration of 800 ms, with the mean values of the aerodynamic and acoustic features, as well as Ps, calculated over the final 50 ms to minimize the presence of transient artifacts.

NN architecture: The baseline model is composed of a multilayer perceptron neural network. The input layer of this network includes seven features: ACFL, MFDR, OQ, SQ, H1-H2, f_0, and SPL, with the output being Ps. Each interconnected hidden layer is equipped with a ReLU activation function and a dropout layer. The hyperparameters of the baseline model were optimized using a 5-fold cross-validation strategy on the synthetic dataset. The iteration of the model that demonstrated the best performance on the validation set across the 5 folds was then selected as the baseline for all subsequent experiments.

The baseline regression model was retrained using transfer learning on laboratory datasets. The optimal performance of the TL strategy was achieved by sequentially freezing the hidden layers. The effectiveness of these models was validated through a subject-independent 10-fold cross-validation.

The baseline model and the TL training employed the Adam optimization algorithm, utilizing mean squared error as the loss function. A learning rate schedule was used, initialized at 0.001. For all experiments, synthetic and laboratory data were min-max normalized.

Results:

Table 1 displays the root mean square error (RMSE) of estimated Ps from ACC data as a function of fine-tuning the NN using TL with sequentially frozen layers. The best performance was achieved by freezing the first layer. This result suggests that the learned weights in the first hidden layer, derived from the voice production model, provide relevant information, thereby enhancing the robustness of Ps estimation from ACC. However, retraining all network parameters without freezing any layers resulted in the loss of the additional variability learned from the numerical voice production model, leading to a higher error in the estimation. Conversely, the poorest performance occurred when retaining all parameters learned from synthetic data (four layers frozen). This finding aligns with our previous work, indicating an existing gap between synthetic and laboratory data, a consequence of the numerical model simplification to simulate the complex human phonatory process.

Table 1. Mean and standard deviation of RMSE from 10-fold cross-validation for subglottal pressure estimated from accelerometer data, applying TL to the NN with sequentially frozen layers.

Number of frozen layers	0	1	2	3
RMSE (cm H_2O)	2.59 ± 0.47	2.30 ± 0.41	2.65 ± 0.56	4.64 ± 0.68

In our preliminary work, the RMSE for subglottal pressure within the same dataset was 2.48 cm H2O. In this proposal, we have reduced this estimation error to 2.30 cm H2O. This reduction represents the best performance reported to date in estimating subglottal pressure from in vivo signals using a single model across multiple subjects. In comparison, the empirically derived formula [2], which computes subglottal pressure using only measurements of SPL and f0, estimates Ps with an RMSE of 2.86 cm H2O. Another NN-based method proposed in [3] reported a mean absolute error (MAE) of 290 Pa in a study involving a single human subject, while our proposal achieved an MAE of 174 Pa in a population of 79 subjects.

Conclusions:

This work presents a methodology based on transfer learning to improve the estimation of subglottal pressure from a neck surface accelerometer, utilizing a framework that integrates machine learning tools with a numerical model of voice production. The validation in laboratory recordings from 79 adult females without voice disorders demonstrates a 7% improvement in the RMSE of subglottal pressure estimation using accelerometer-based features, compared to our preliminary findings. These results suggest that the triangular body-cover model offers a good general representation of typical sustained phonation across a wide range of conditions, enhancing the robustness of the non-linear mapping between ACC-based features and Ps. These outcomes strengthen our ongoing efforts to combine a numerical voice production model with machine learning tools, providing a viable approach for physiologically relevant measurements in both laboratory and ambulatory settings for the assessment of vocal function. While this method is initially described in the context of subglottal pressure, its applicability can be extended to the estimation of muscle activation, vocal fold contact pressure, and other measures. However, obtaining clinical recordings for validating these additional features is challenging.

Acknowledgements:

This research was supported by the National Institutes of Health (NIH) National Institute on Deafness and Other Communication Disorders grant P50 DC015446; by ANID grants BASAL FB0008, FONDECYT 1230828, and Beca de Doctorado Nacional 21190074. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

References:

[1] Ibarra, E., Parra, J., Alzamendi, G., Cortés, J., Espinoza, V., Mehta, D., Hillman, R., and Zañartu, M. (2021). “Estimation of subglottal pressure, vocal fold collision pressure, and intrinsic laryngeal muscle activation from neck-surface vibration using a neural network framework and a voice production model”. Frontiers in Physiology, section Computational Physiology and Medicine, 12, 1419.

[2] Titze I., Svec, J., and Popolo P. (2003) “Vocal dose measures: quantifying accumulated vibration exposure in vocal fold tissues”. J Speech Lang Hear Res. 46(4): 919-32.

[3] Zhaoyan, Z. (2022). “Estimating subglottal pressure and vocal fold adduction from the produced voice in a singlesubject study (L)”. J. Acoust. Soc. Am., 151 (2): 1337–1340.

A DEEP LEARNING BASED TWO-STEP APPROACH FOR THE DETECTION OF SPECIFIC TYPES OF PIG PHONATION

Patrick Schlegel[1], Jennifer Long[2]

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

***Keywords*: e.g. Voice; Pigs; Deep learning; Audio processing**

Abstract:

Introduction:

To produce a clear and healthy voice, the vocal folds (VF) located in the larynx need to oscillate regularly and symmetrically. However, greater injury to the VF tissue can lead to scarring during the healing process, permanently changing VF physiology and hence distorting VF oscillations, leading to an adverse and persistent voice change [1]. Therefore, treatments to prevent scarring or heal existing scars are of interest. One such potential treatment that is currently under development is a stem cell-based outer vocal fold replacement (COVR) that is in the animal testing stage. To assess the effect of COVR on injured pig vocal folds, structural and functional recovery of the pig voice after COVR implantation are of interest.

The assessment of functional recovery is especially challenging, as pigs cannot be instructed to phonate in specific ways and only the higher pitched type of pig phonation (here called “squeals”) that involves the VFs is of interest. Lower pitched phonation (here called “grunts”) is theorized to be of non-laryngeal origin [2] and hence may not be affected by VF injury. Further, if more stressful conditions in which the pig is actively stimulated to squeal are avoided, recording sessions need to be prolonged to capture spontaneous pig squealing and will inevitably include large amounts of data in which the pig does not squeal or does not phonate at all. Therefore, from a large amount of recorded acoustic data, only a small fraction can be utilized [3]. For the automatized extraction of squeal-like pig-phonation we propose a deep learning-based two-step approach.

Methods:

For this work a dataset consisting of about 1500 hours of pig phonation of 11 pigs is utilized. The data was originally collected for a previous publication [3]. The data were partially labeled in a double-blind non-exhaustive process, with approximately 0.4% of the data identified as squeals. As the line between “grunts” and “squeals” can be blurry and annotators did not label all the data, it is to be expected that the actual squeal content of the data is significantly higher. Further, large sections of the data consist of relative silence or noises very distinctly different from pig phonation. Lastly, pig squeals are expected to differ between recordings before and briefly after surgery, with post-surgery squeals sounding more grunt-like but still distinct from grunts.

To factor in all these different properties of the data, we employ a two-step approach as depicted in Figure 1. The preprocessing is done similar to [4]. First, Mel-spectrograms are calculated for multiple hour-long segments of acoustic data with the number of FFTs being set to 2048 and the number of Mel-filters being set to 128. The resulting Mel-spectrograms are then divided into 128×64 windows with 50% overlap. This corresponds to a length of approximately 1.25s in the time domain which corresponds to the average length of a labeled squeal after removing outlier labels. Mel-spectrogram segments are labeled with values between 0 and 1, representing the percentage of the respective window being identified as a squeal in the time domain. Mel-spectrogram windows and labels are randomized and divided into training, validation, and test data set (60%/20%/20%).

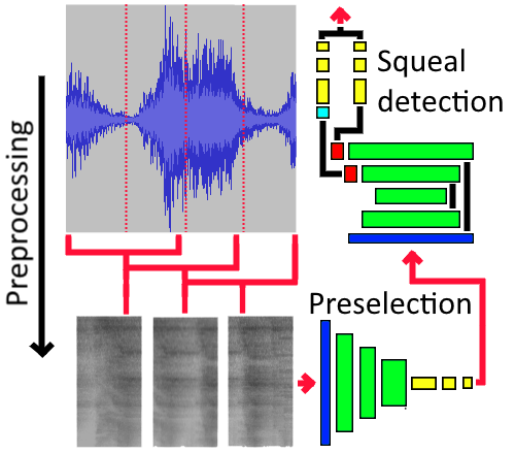


Figure 1: Summary of the squeal detection process

In the first detection step the data is fed into a network based on the Le-net architecture [5] sampling from both classes (contains a squeal (fraction) and contains no squeal) equally. Instead of simple convolutional layers the network features residual blocks as in [4] to increase depth. To exclude as little true squeals as possible in this first step a weighted cross entropy loss is used for training, overweighting the squeal-class five-fold. Different types of data augmentation, batch normalization and drop-out layers are added to increase regularization and avoid over-fitting.

In the second detection step all remaining samples are fed into the second network based on the work of Ding et al [4]. As they focused on the detection of multiple different classes detecting exact onset and offset positions of the different acoustic events, and we are only working with a simpler two-class problem, this second network is a simplified version of the one proposed by Ding et al. It features an hourglass model with only two instead of four layers consisting of only single residual blocks that respectively feed into two branches with gated recurrent units (GRU)s. These branches then "vote" on the classification of a sample. In Table 1 important Hyperparameters for both models are given.

Table 1: Important Hyperparameters

Parameter	Model 1	Model 2
Learning rate	1E-6	5E-6
Epochs	50	50
Image height	128	128
Image width	64	64
Batch size	64	8
Weights	0.2 ,1.0	0.25, 1.0

Results and discussion:

In Figure 2 false positive rate (FPR), Brier Score, true positive rate (TPR) and soft true positive rate (STPR) are visualized. For the calculation of TPR all samples that contained any amount of squeal were categorized as belonging to the squeal-class. In contrast, for STPR this was only the case for samples consisting of at least 40% squeal. As can be seen in the figure, the first model, and to a lesser extent the second one, overfit on small squeal fragments that occur when one sample only contains the very beginning or end of a squeal, but generalize well for all samples that contain larger squeal sections and not-squeal samples. Good results were achieved with very low learning rates, final FPR, Brier score, TPR and STPR on the testing set were 0.11, 0.08, 0.90 and 0.99 for model 1 and 0.37, 0.25, 0.86 and 0.92 for model 2. Due to the high number of not labeled squeals STPR remains high. A comparative evaluation of these remaining false positives and true positive squeals did not yield any audible difference. Squeals are detected before and after surgery with overall comparable performance.

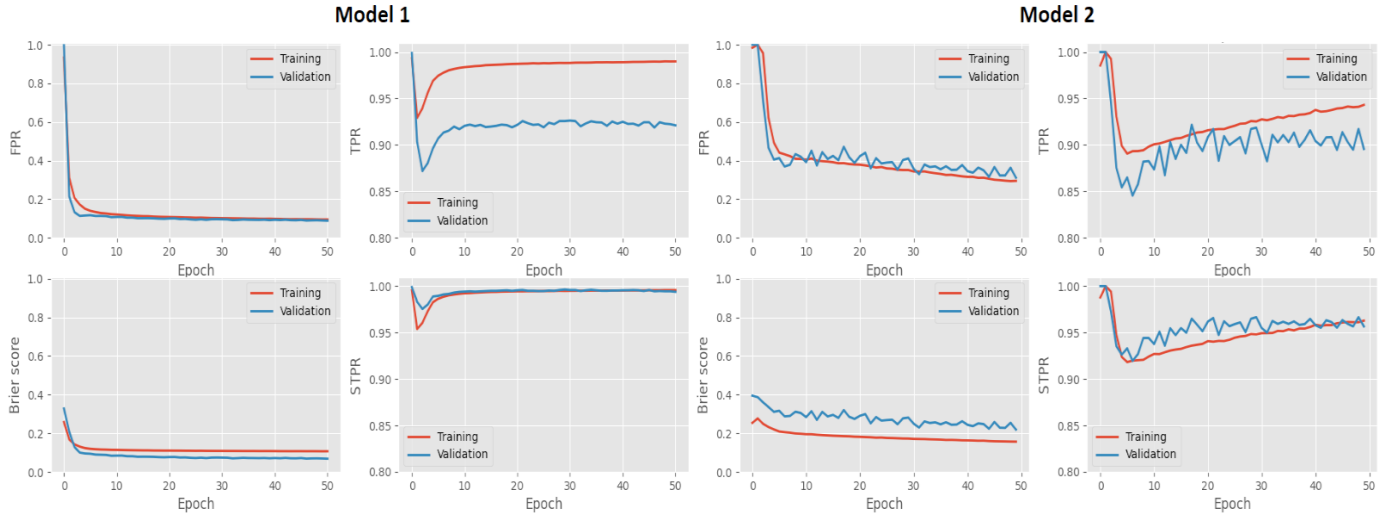


Figure 2: FPR, Brier Score, TPR and STPR for training and validation datasets plotted across the training epochs for the two models.

Conclusions:

The described pipeline is suitable for automatic detection of pig squeals within the specific settings the training data was recorded in, however, some retraining will be necessary if these settings change. Whilst pig squeals of a minimum length of 0.5s in the time domain can be identified reliably, a modification of the second model may also allow for the exact detection of event onset and offset, similar to the work of Ding et al. [4].

Acknowledgements:

VA Office of Research and Development MERIT I01 RX003649 (Long)

References:

[1] Hansen, J.K., Thibeault, S.L. (2006) "Current Understanding and Review of the Literature: Vocal Fold Scarring." Journal of Voice 20(1), 110-120.

[2] Riley, J.L., Riley, W.D., Carroll, L.M. (2016) "Frequency Characteristics in Animal Species Typically Used in Laryngeal Research: An Exploratory Investigation." J. Voice 30, 767.e17–767.e24.

[3] Schlegel, P., Yan, K., Upadhyaya, S., Buyens, W., Wong, K., et al. (2023) "Tissue-engineered vocal fold replacement in swine: Methods for functional and structural analysis." PLOS ONE 18(4): e0284135.

[4] Ding, W., He, L. (2020) "Adaptive Multi-Scale Detection of Acoustic Events." IEEE/ACM Transactions on Audio, Speech, and Language Processing 28, 294-306.

[5] LeCun, Yann, et al. "Backpropagation applied to handwritten zip code recognition." Neural computation 1.4 (1989): 541-551

Multi-Level Label Hierarchy for Disordered Voice Databases

Rijul Gupta[1], Dhanshree R Gunjawate[2], Dharshini Manoharan[2], Duy Duong Nguyen[2], Sarah Emmett[2], Daniel Novakovic[2], Craig Jin[1], Catherine Madill[2]

[1] Computing and Audio Research Laboratory, School of Electrical and Computer Engineering, The University of Sydney, Australia

[2] Voice Research Laboratory, Sydney School of Health Sciences, Faculty of Medicine and Health, The University of Sydney, Australia

***Keywords*: e.g. Voice Disorder Recognition; Machine Learning; Diagnosis Classification; Data Quality Improvement**

Abstract:

Objectives / Introduction:

Free and publicly available databases have been an impetus for substantial progress in Machine Learning (ML) over the past two decades in various domains including Computer Vision, Natural Language Processing and Speech Processing. However, in Voice Disorder Detection/Recognition, such databases' availability has been limited. Three databases namely, Saarbruecken Voice Database (SVD) [1], VOice ICar fEDerico II (VOICED) [2], and TORGO [3] are publicly and freely available but the amount of data available within these databases is substantially limited. SVD contains 2042 sessions and 71 diagnostic labels (classes), VOICED contains 208 sessions and 24 classes, and TORGO contains 30 sessions and 2 classes.

Current ML systems cannot train when data availability is low, and the number of classes is high, as in the available databases. Due to this limitation, researchers have often resorted to performing binary classification (healthy vs pathological) or combining pathologies under arbitrary sub-classifications e.g. hypokinetic vs hyperkinetic. However, these classifications are often clinically inappropriate, irrelevant or have no consensus across the clinical community.

We address this problem by re-classifying labels in clinically relevant categories using a consensus process across a range of clinicians and present a multi-level label hierarchy that can be used for collating data into clinically relevant classification labels.

Methods:

The data from SVD, VOICED and TORGO was collected, and their diagnosis was extracted using Python scripts. As the diagnosis in SVD were available in German, we used Google Translate to translate them to English. Duplicates from the resulting set of diagnosis were removed by manual inspection of the list, and similar labels were grouped together into one class e.g. GERD and gastric reflux. This final set of clean labels from the databases was discussed with the author clinicians and we identified that there was a potential lack of consensus in how to classify these labels. We addressed the lack of consensus in three stages: (1) Two focus group sessions were held to establish a classification system based on published frameworks [4]. (2) A Qualtrics [5] survey where seven clinicians, including Otorhinolaryngologists and Speech-Language Pathologists, classified the input labels from the databases according to the preliminary classifications. (3) Follow up focus group sessions were held with clinicians and the engineering team to establish a multi-level classification system. The results of the survey were analyzed using Microsoft Excel and Python scripts. After analysis a hierarchy was established using percentage of votes from the survey as confidence rating of re-classification category. While a patient may be diagnosed with multiple disorders and have co-factors, the current research focus is on single diagnosis classification, with the intention of leaving multiple disorders and co-factors as a future and active research area.

Results:

We grouped the 103 diagnostic labels available in the databases into 4 levels of classification containing 3 classes at level 0, 4 classes at level 1, 4 classes at level 2 and 95 classes at level 3 as shown in Fig 1.

52 labels from the available dataset were either classified as not being a diagnosis or had no consensus. Out of these, muscle tension dysphonia was most often confused with other pathologies, and subgroups of organic i.e. neuro muscular, structural and trauma were often confused with one another. This result is summarized in Table 1. We are curently working on a new classification strategy that better encompasses these cases and should result in more consensus.

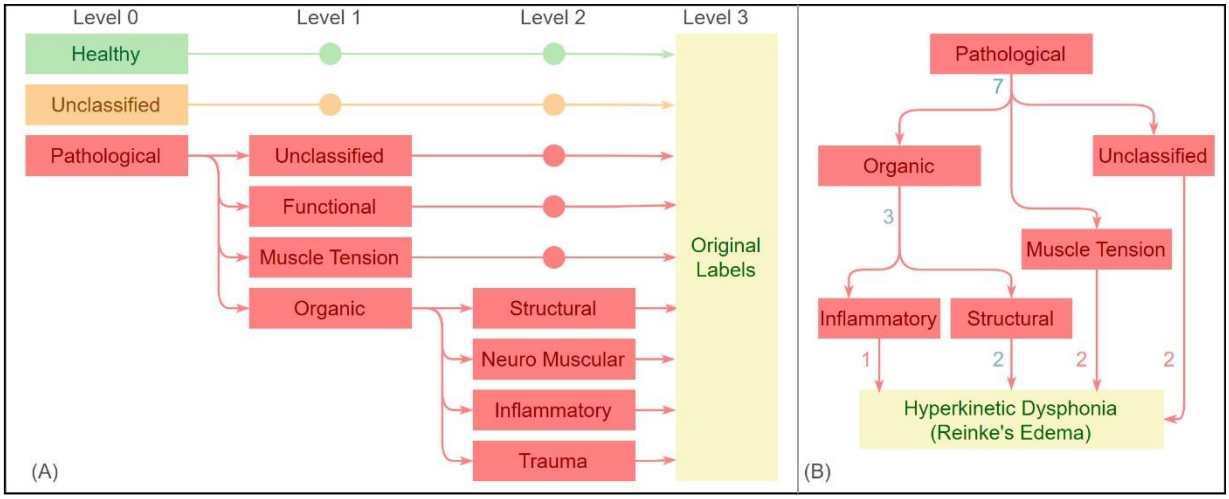


Figure 1 (A) shows the ideal hierarchy designed from top (level 0) to bottom (level 3) the dots on level 1 and level 2 indicate that the classification from previous level is carried forward to the current level if further sub-classification is not available. (B) an example of poor consensus on classifying a particular input label. We note here that Organic > Inflammatory got 1 vote, and Organic > Structural, Muscle Tension got 2 votes and 2 clinicians voted that this is not a (proper) diagnosis. Using the hierarchy and votes we classify the sample as Pathological > Organic > Structural > Hyperkinetic Dysphonia (Reinke's Edema)

	P>U	F	MT	O > I	O > NM	O > S	O > T
P>U	71	5	21	8	20	35	2
F	5	8	1	-	2	1	-
MT	21	1	21	1	9	13	-
O>I	8	-	1	11	-	9	1
O>NM	20	2	9	-	25	6	-
O>S	35	1	13	9	6	47	2
O>T	2	-	-	1	-	2	3

Table 1 Highlighting how often different pathologies are confused with others calculated as unique pathologies where a pathology is confused with another. The left diagonal axis indicates the number of times a diagnosis appears in the classification set. Keys: {P>U=Pathological > Unclassified, F=Functional, MT=Muscle Tension, O>I=Organic Inflammatory, O>NM=Organic Neuro Muscular, O>S=Organic Structural and O>T=Organic Traumatic}

Conclusions:

Through this research effort we were able to group the 103 diagnostic labels available in the databases into 4 levels of classification that can now be used by machine learning researchers to produce more clinically relevant results This work will soon be released as open-source software as part of a disordered voice recognition benchmark.

Acknowledgements:

We would like to acknowledge all the clinicians that were involved in re-classifying the labels obtained from the databases. We would also like to acknowledge Doctor Liang Voice Program and Faculty of Engineering for funding this research.

References:

[1] B. Woldert-Jokisz, "Saarbruecken voice database." 2007.

[2] U. Cesari, G. De Pietro, E. Marciano, C. Niri, G. Sannino, and L. Verde, "A new database of healthy and pathological voices," *Comput. Electr. Eng.*, vol. 68, pp. 310–321, May 2018, doi: 10.1016/j.compeleceng.2018.04.008.

[3] F. Rudzicz, A. K. Namasivayam, and T. Wolff, "The TORGO database of acoustic and articulatory speech from speakers with dysarthria," *Lang. Resour. Eval.*, vol. 46, no. 4, pp. 523–541, Dec. 2012, doi: 10.1007/s10579-0119145-0.

[4] C. L. Payten, G. Chiapello, K. A. Weir, and C. J. Madill, "Frameworks, Terminology and Definitions Used for the Classification of Voice Disorders: A Scoping Review," *J. Voice*, p. S089219972200039X, Mar. 2022, doi: 10.1016/j.jvoice.2022.02.009.

[5] "Qualtrics." Qualtrics, 2005. [Online]. Available: https://www.qualtrics.com

DEEP LEARNING-BASED LARYNGOTRACHEAL AND AIRWAY ANATOMICAL LANDMARK DETECTION FOR ULTRASOUND-GUIDED INTERVENTIONS

Gregory R. Dion, MD, MS, FACS[1], Laura Brattain, PhD[2]

[1] Department of Otolaryngology – Head and Neck Surgery, University of Cincinnati, Cincinnati, OH, USA
[2] College of Medicine, University of Central Florida, Orlando, FL, USA

***Keywords*: Deep Learning, Machine Learning, Airway, Larynx**

Abstract:

Objectives / Introduction: Laryngotracheal visualization is necessary for diagnosis of vocal fold mobility disorders, delivery of therapeutics such as onabotulinum toxin for laryngeal dystonia or hyaluronic acid for vocal fold augmentation, and airway access for delivery of serial intralesional steroid injections for subglottic stenosis or percutaneous tracheotomy. To date, few modalities outside of flexible laryngoscopy and bronchoscopy exist for these interventions. Similar to the movement of laryngeal procedures from the operating room to the clinic, over the past few decades, percutaneous tracheotomy has grown in popularity as it can be readily performed at the bedside by an intensivist, obviating the requirements of a surgeon and/or valuable operating room time. Since there are critical anatomical structures in the vicinity of the larynx and trachea, ultrasound guidance can help improve the accuracy and safety of tube placement or therapeutic delivery. The goal of this work is to demonstrate the feasibility and accuracy of deep learning-based neck anatomical landmark detection for assisting laryngotracheal procedures.

Methods: We collected neck ultrasound data on 25 healthy subjects using the Terason uSmart 3200 (Teratech, Burlington, MA) and linear probe 15L4A with the neck imaging preset. Imaging depth was set at 3 cm. Data collections were approved by the Institutional Review Board at Massachusetts Institute of Technology and University of Cincinnati. For each subject, we recorded multiple cineloops of 10-second length each. The clips mimic how experts use ultrasound to identify the optimal tracheal space for needle insertion. The ultrasound probe was in short-axis and the scans included views of the anterior of the neck from hyoid bone to suprasternal notch. Ultrasound frames from the cineloops were extracted and annotated by two domain experts. Key anatomical landmarks such as thyroid cartilage (TC), cricoid cartilage (CC), thyroid gland (TG), strap muscles (SM), and tracheal rings (TR) were identified with bounding boxes using our custom tool for AI development.

An AI model was trained by using the You-Only-Look-Once (YOLO) v7 real-time object detection deep learning network to detect TC, SM, TG, and TR in any given ultrasound frame. The YOLO model was pretrained with the publicly available MSCOCO dataset and finetuned with our neck ultrasound data (Figure 1). Stochastic Gradient Descent along with a linear warmup of the learning rate from 0.01 to 0.1 followed by a cosine decay is used to optimize the model. A number of image augmentations were used to build invariance / robustness into the model including saturation, intensity, reflections about the vertical axis, and random cropping. Augmentations such as rotation or reflections about the horizontal axis were not included as they would alter the image outside the distribution of data.

Results: Total of 28,000 frames were used for the training. Data were split into test/validation/train by subject with 3fold cross validation. Precision-recall curves for TC, SM, TG, and TR were computed. AUCs are 0.74, 0.92, 0.95, and 0.84, respectively (Figure 2). Average AUC for all classes is 0.86. The inference speed on a GPU-enable Terason tablet is at least 30 frames per second, indicating real-time performance. The average F1 score for the above classes are 0.483, 0.719, 0.763, and 0.587, respectively, with an all class average F1 score of 0.638.

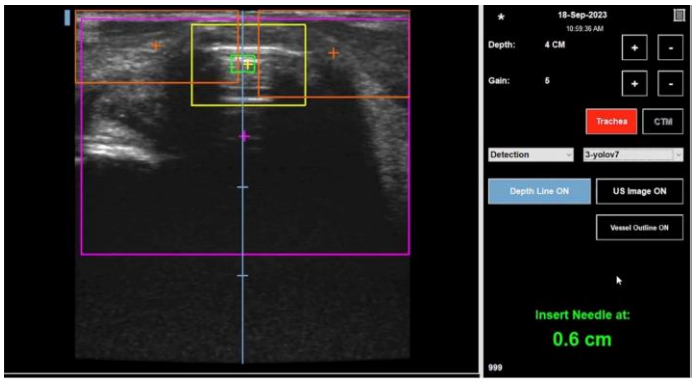


Figure 1. Example AI model detection and classification. A YOLO model detects: strap muscles (orange), thyroid glands (purple), and a tracheal ring (yellow).

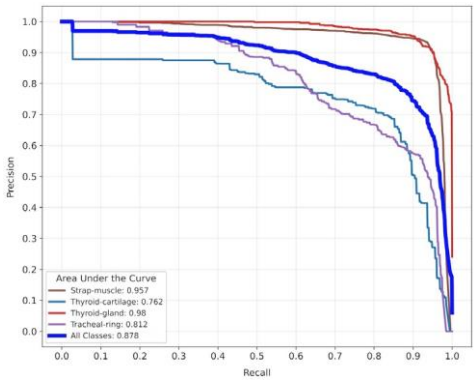


Figure 2. Precision-recall curves for thyroid cartilage, strap muscles, thyroid gland, and tracheal ring. AUCs are 0.74, 0.92, 0.95, and 0.84 , respectively. Average AUC is 0.86.

Conclusions: A deep learning-based AI algorithm has been developed to detect key anatomical landmarks along the anterior of the neck for future autonomous interventions. The next step is to further improve the robustness by training on much more diverse populations of varying neck sizes and body builds. The algorithm is being integrated into an airway access robotic assistant. It has the potential to expand into laryngeal and laryngotracheal interventions to include vocal fold augmentation, therapeutic delivery, and tissue biopsy in a rapid, safe procedure accessible with less required technical acumen.

AUTOMATIC DETECTION AND MOTION TRACKING OF THYROID AND ARYTENOID CARTILAGES IN DYNAMIC TRANSLARYNGEAL ULTRASOUND: A FEASIBILITY STUDY FOR THE DIAGNOSIS OF VOCAL FOLD PARALYSIS

Trung Kien Bui[1], Muriel Lefort[2], Agnès Rouxel[3], Hervé Guillemet[4], Juliette Dindart[1], Christophe Nioche[1], Crystal Lin[5], Christophe Trésallet[6], Frédérique Frouin[1]

[1] LITO, Inserm, Institut Curie, Orsay, France
[2] LIB, Sorbonne Université, Paris, France
[3] Department of Nuclear Medicine, AP-HP, Bobigny, France
[4] Apteryx, Gif-sur-Yvette, France
[5] Mindray, Créteil, France
[6] Department of Surgery, AP-HP, Bobigny, France

***Keywords*: Vocal fold paralysis; Arytenoid tracking; Ultrasound imaging; Deep Learning.**

Abstract:

Objectives / Introduction:

Recurrent nerve lesions and, thus, vocal fold paralysis (VFP) are some of the major risks (occurrence of 3-5%) associated with neck surgery. The reference diagnostic technique for this lesion is laryngoscopy, a minimally invasive procedure. Dynamic TransLaryngeal UltraSound (dTLUS) acquired during free breathing was proposed as a noninvasive alternative to assess vocal fold motion and detect possible paralysis [1]. Previous works of our research group have shown encouraging performances of dTLUS to diagnose VFP after thyroid or parathyroid surgery by applying this technique to the laryngeal prominence of the thyroid cartilage, including the arytenoids in the axial view [2]. VFP was characterised by quantitative measures of the laryngeal tract motion using three landmarks (thyroid, left and right arytenoid cartilages) on two images of dTLUS corresponding to the start and end of an abduction or adduction phases [2], [3]. To improve this measure, we now aim to track these laryngeal markers on the whole set of images in the video. The main challenge is to robustly detect the three landmarks, regardless of the quality of the dTLUS data. The current study uses YOLO [4], a deep learning algorithm based on convolution neural networks, to detect these markers in each video frame and track their motion. This automatic detection and motion tracking could allow the assessment of laryngeal mobility and VFP diagnosis.

Methods:

This study utilised two retrospective databases: a training database of 197 subjects and a test database of 68 patients. Both databases comprised dTLUS scans acquired with 5 different US devices. Each dTLUS acquisition lasted between 10 and 30 seconds (at a frame rate of 30 images per second) while the subject was asked to breathe freely. An expert manually selected between 2 and 16 images within each dTLUS scan, mainly representing extreme positions of the glottis in the abduction and adduction. Next, she pointed to the centres of the thyroid and both arytenoid cartilages (left and right) and defined a fixed-size bounding box centred on the selected points. A total of 1280 and 415 images were respectively labelled for the training and test databases. Then, the YOLOv7-Tiny model [4] was trained from scratch for 2 object classes*, 'thyroid'* and *'arytenoid',* on the first dataset using the default hyperparameters and 300 epochs. We evaluated its performance using the second dataset with the standard *intersection over union* (IoU) threshold metric to quantify the matching between a prediction and the ground truth. Due to the consistent size of ground-truth bounding boxes and the variability in the size of YOLO-detected bounding boxes, we proposed calculating the distance between the centres of each detected structure and its corresponding ground-truth location. We set a threshold of 0.04 using normalised coordinates of the images to define accurate detections. We eventually applied the trained model to the whole set of dTLUS acquisitions, and the displacement of detected points was color-encoded to illustrate their motion across video sub-sequences, defined as an abduction or an adduction phase.

Results:

On the test set, using the YOLO default confidence threshold of 0.001 and an IoU threshold of 0.65, we obtained a precision of 0.622 and a recall of 0.654. The mean average precision at an IoU of 0.5 was 0.537. Additionally, the confusion matrix proposed by YOLO in Figure 1 shows that the model correctly predicts 75% (311 out of 415) of thyroid cartilage locations and 70% (580 out of 830) for both arytenoid cartilage locations. *Figure 2a* illustrates the ability of the trained model, to detect the three landmarks (thyroid cartilage, left and right arytenoid cartilages) in an ultrasound image.

By using the distance between box centres instead of IoU on the test set, the trained model demonstrated an overall detection rate of 99.8% (414 out of 415 images) for at least one of the three and a detection rate for the three structures equal to 64.3% (267 out of 415 images). The thyroid cartilage was correctly identified in 81% of images, and the left and right arytenoid cartilages in 82% and 86% of the images. In most cases of detection failure, the algorithm was able to identify the landmark on another image of the same dTLUS acquisition. More precisely, thyroid cartilage was not detected in at least one of the labelled images of the dTLUS in 35 patients (image detection level), and it was not detected in all the labelled images of the same dTLUS in 7 patients (global detection level). For the left arytenoid cartilage, the failure at the image detection level corresponded to 36 patients and at the global detection level to 8 patients. For the right cartilage, there were 27 and 5 patients, respectively. These results encouraged us to use the YOLO model for tracking the landmarks in the whole set of dTLUS scans. To allow an interpretation of the displacement of landmarks, we restrict the tracking to abduction motion (*Figure 2b*) or adduction motion (*Figure 2c*). Moreover, we consider the motion of the arytenoid cartilages relative to the thyroid cartilage, thus partly correcting for probe or patient movement during the acquisition. The displacement of the two arytenoids over time shown in *Figure 2b* aligns with the normal vocal fold motion observed in a healthy subject, with a symmetrical large amplitude motion. In contrast, *Figure 2c* reveals a reduced displacement of the left arytenoid cartilage (on the right side of the figure), while the right one shows normal movement. This asymmetry was consistent with unilateral left vocal fold paralysis, as confirmed by laryngoscopy.

Figure 1: The confusion matrix on the test set using a confidence threshold of 0.001 and an IoU threshold of 0.65.

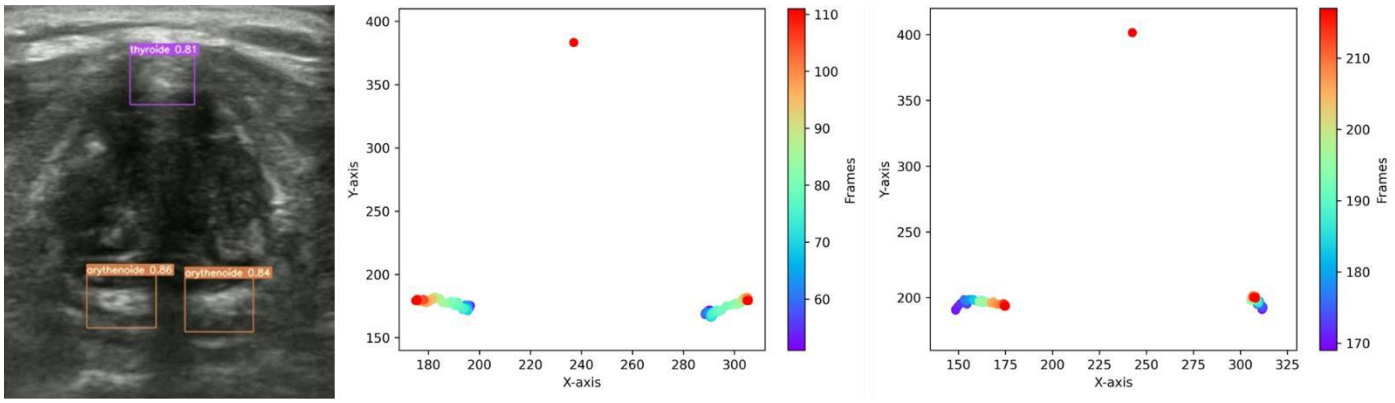


Figure 2: (a) Two arytenoid cartilages (orange boxes) and a thyroid cartilage (purple box) detected in a single frame. Time color encoded position of the left and right arytenoid cartilages in a healthy subject during an abduction phase (b) and in a subject with unilateral vocal fold paralysis (left arytenoid, on the right side of the figure) during an adduction phase (c).

Conclusions:

Based on the evaluation metrics, we demonstrate the ability of the YOLO model to effectively detect thyroid and arytenoid cartilages from dTLUS images, enabling accurate tracking of arytenoid cartilage movement. We observed limitations in some specific cases, including duplicate and missed detections. The model needs further refinement to enhance its generalisability through data augmentation and optimisation. We recently provided experts with a system for video labelling, which will be applied to our training data to improve the model's ability to track movement patterns. Furthermore, we will examine quantitative indices derived from motion tracking.

Acknowledgements:

The authors thank ANR for its financial support (grant number ANR-22-CE19-0035) and Lydia Abdemeziem for her first tests to detect thyroid and arytenoid cartilages using the YOLOv4 model.

References:

[1] M. Dedecjus, Z. Adamczewski, J. Brzeziński, and A. Lewiński, 'Real-time, high-resolution ultrasonography of the vocal folds—a prospective pilot study in patients before and after thyroidectomy', *Langenbecks Arch. Surg.*, vol. 395, no. 7, Art. no. 7, Sep. 2010, doi: 10.1007/s00423010-0694-2.

[2] D. S. Lazard *et al.*, 'Transcutaneous Laryngeal Ultrasonography for Laryngeal Immobility Diagnosis in Patients with Voice Disorders After Thyroid/Parathyroid Surgery', *World J. Surg.*, vol. 42, no. 7, Art. no. 7, Jul. 2018, doi: 10.1007/s00268-017-4428-2.

[3] H. Bergeret-Cassagne *et al.*, 'Sonographic Dynamic Description of the Laryngeal Tract: Definition of Quantitative Measures to Characterize Vocal Fold Motion and Estimation of Their Normal Values: Sonographic Dynamic Description of the Laryngeal Tract', *J. Ultrasound Med.*, vol. 36, no. 5, Art. no. 5, May 2017, doi: 10.7863/ultra.16.05014.

[4] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, 'YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors' arXiv, Jul. 06, 2022. doi: 10.48550/arXiv.2207.02696.

DO SPEECH TASKS INFLUENCE MACHINE LEARNING PERFORMANCE IN DYSPHONIA DETECTION?

Ahmed M Yousef[1], Adrián Castillo-Allendes[1,2], Eric J Hunter[1]

[1] Department of Communication Sciences and Disorders, University of Iowa, Iowa City, Iowa, USA
[2] Department of Communicative Sciences and Disorders, Michigan State University, East Lansing, Michigan, USA

***Keywords*: Voice Disorders; Voice Assessment; Machine Learning; Speech Acoustics**

Abstract:

Objectives / Introduction: Machine learning, despite its potential as a promising objective tool, could be applied to voice assessment methods which are inherently subjective and highly variable. Yet machine learning tools remain in their infancy in applications towards detecting voice disorders from speech and vocal tasks. The present work examines the impact of different speech recording variants on the performance of machine learning models in detecting dysphonia, aiming to determine the most effective task for optimal outcome.

Methods: Voice samples were recorded from 200 participants: 148 individuals with different dysphonia etiologies and 52 control group participants. Acoustic metrics and quantities (including perturbation, noise, cepstral, and spectral analyses), were estimated from the speech recordings using a MATLAB script integrated with Praat. Metrics were extracted from multiple types of speech productions: a sustained vowel /a:/, connected speech, concatenated samples (a combination of vowel and speech recordings), as well as other variants of the recordings (e.g. various lengths). Machine learning models were developed from extracted acoustic metrics, and evaluated for accuracy, sensitivity, and specificity to compare the impact of different speech recording variants on dysphonic voice classification.

Results and Conclusions: The performed analysis enabled accurate extraction of different acoustic measurements, illustrating how machine learning techniques could be implemented to support voice analysis and facilitate successful classification of dysphonic versus non-dysphonic speakers. The comparative analysis of the different models demonstrated discrepancies in the classification accuracy among the different speech samples. Indications of an optimal speech task or sample length that yield the best machine learning performance allow for more efficient machine learning models utilized for diagnosing voice disorders and tracking voice change. The outcome of this work is a vital step toward creating efficient processes – paving the way for accurate AI-driven tools in the screening of voice disorders and the enhancement of clinical voice assessment.

Classification of vocal fold paralysis versus controls based on voice recordings with a Random Forest model: robustness validated using an external database

Juliette Dindart[1], Hélène Massis[2], Trung Kien Bui[1], Agnès Rouxel[1,3], Claire Pillot-Loiseau[2], Frédérique Frouin[1]

[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] Department of Nuclear Medicine, AP-HP, Bobigny, France

***Keywords*: Vocal fold paralysis; Vocal recordings; ComBat harmonization; Machine learning**

Abstract:

Introduction:

Cervical surgery is associated to a risk of vocal fold paralysis (VFP) at a rate of 3 to 5% of patients in case of recurrent nerve damage. To assess vocal fold paralysis, laryngoscopy is the commonly used method. However, this minimally invasive technique is unpleasant for the patient, and it is not always applied in standard care. This can result in missing or delaying the diagnosis of VFP. An alternative method using vocal recordings could help to characterize surgically induced dysphonia. Vocal parameters of jitter, shimmer and harmonic to noise ratio (HNR) are known to be modified in patients with recurrent nerve lesion due to surgery [1]. Many classification approaches to define controls and pathological voices have recently been proposed based on deep learning or on machine learning approaches [2]. Most of them were based on publicly available databases, such as the Saarbruecken Voice Database (SVD) [3], and/or the Advanced Voice Function Assessment Database (AVFAD) [4]. Our goal was to create a robust classification model to distinguish between VFP and control subjects. To reduce the black-box effect of artificial intelligence methods, we aim at building a simple and interpretable model and testing its relevance in a different context, with different recording conditions and speakers with different mother tongues.

Methods:

We first used the German SVD database, consisting of vocal recordings of vowels and short sentences. We selected recordings from control subjects (428 females and 259 males) and from subjects afflicted with VFP (139 females and 74 males). For each subject, we selected continuous vocalized /a vowel, produced at a normal pitch. The voice parameters were then extracted from the recordings using the Parselmouth library (Python version of Praat). A total of 14 parameters were extracted from the recordings. These vocal features were reduced to 5: fundamental frequency (f0), standard deviation of f0, local jitter, local shimmer, and Harmonic To Noise ratio (HNR), to avoid large redundancies with highly correlated parameters. Using these five complementary features, we trained a Random Forest classifier with a stratified five-fold cross validation based on the optimization of the accuracy, using Scikit-Learn library. To test the robustness of the model, we used an independent dataset. We extracted from the Portuguese database AVFAD 396 recordings (363 control subjects and 33 patients with VFP, vocalized /a vowel, produced at a normal pitch). To overcome the differences between the five parameters of interest that exist between the two databases, we applied the ComBat data harmonisation [5] (Figure 1). The 5 models, one per fold of the cross-validation, were tested on the AVFAD database, before and after harmonization.

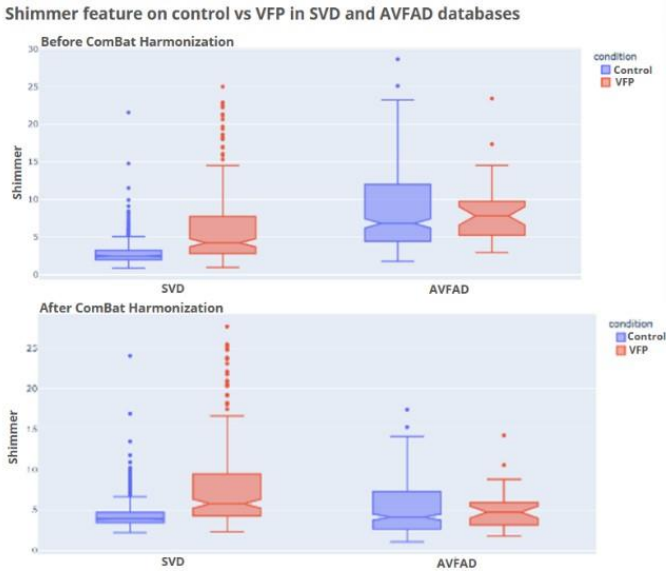


Figure 1: Local shimmer before and after Cpmbat harmonization between SVD and AVFAD databases.

Results:

Using the five-fold cross validation, the five Random Forest models provided a mean accuracy (standard deviation) of 0.85 0.01. The feature importance is shown in Figure 2, highlighting the impact of local shimmer, local jitter, and HNR for the classification, while the mean f0 has a reduced role in the model. When applied the models to the data extracted from the AVFAD recordings, we obtained a mean accuracy between the 5 folds of 0.61 0.09, showing a poor exportability of the models. After investigating these underwhelming results, we uncovered significant differences in the five parameters of interest, especially for HNR, and local shimmer between the SVD and the AVFAD (see Figure 1 for local shimmer). If the variations between normal and VFP values were as expected for both databases, there was a clear shift of values between the two databases. These shifts clearly explain that the classification results could not be satisfying on the AVFAD database. The ComBat procedure was thus applied to the five features to harmonize them between the two databases. After the harmonization, the new Random Forest models provided the same mean accuracy between the 5 folds on the SVD database (0.84 0.01) than before harmonization. When applied to the external AVFAD database, these new models provided a mean accuracy of 0.81 0.02, demonstrating the high robustness of our approach.

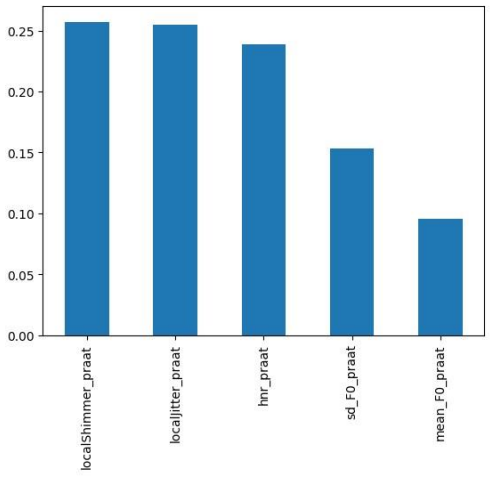


Figure 2 : Feature importance for the classification.

Conclusions:

It is well known that the values of local jitter, local shimmer and HNR are sensitive to variations in recording conditions [6]. By using an external database, we demonstrated the robustness of a simple Random Forest model based on 5 vocal features in accurately categorizing recordings between control and paralyzed subjects, provided that vocal features are harmonized between databases. This harmonization step enables to take advantage of the different recording conditions and of the mother tongues. As the model was built with recordings of /a vowel phonation, future work will aim at improving the classification performance, incorporating other vowels recordings (/i and /u), while maintaining a robust and interpretable model.

Acknowledgements:

The authors thank ANR Vocalise (grant number ANR-22-CE19-0035) and the Paris-Saclay University for their financial support.

References:

[1] Dejonckere P, Bradley P, Clemente P, et al. (2001). A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques. *European Archives of Oto-Rhino-Laryngology* 258, 77–82. doi:10.1007/s004050000299.

[2] Ribas D, Pastor MA, Miguel A, et al. (2023). Automatic Voice Disorder Detection Using Self-Supervised Representations. *IEEE Access*, 11, 14915-14927. doi:10.1109/ACCESS.2023.3243986.

[3] Barry WJ, Pützer M. (2007) Saarbrücken Voice Database, Institute of Phonetics, Univ. of Saarland, http://www.stimmdatenbank.coli.uni-saarland.de/

[4] Jesus LMT, 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. doi:10.5772/intechopen.69643.

[5] Behdenna A, Haziza J, Azencott CA, et al. (2020). pyComBat, a Python tool for batch effects correction in highthroughput molecular data using empirical Bayes methods. bioRxiv doi:10.1101/2020.03.17.995431.

[6] Doherty ET, Shipp T (1988). Tape Recorder Effects on Jitter and Shimmer Extraction. *Journal of Speech Language and Hearing Research,* 31(3), 485–490. doi:10.1044/jshr.3103.485.

OBJECTIVE ASSESSMENT OF FUNCTIONAL DYSPHONIA BASED ON MACHINE LEARNING AND HIGH-SPEED VIDEOENDOSCOPY

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

[1] Division of Phoniatrics and Pediatric Audiology, Department of Otorhinolaryngology, Head and Neck Surgery, University Hospital Erlangen, Friedrich-Alexander-University Erlangen-Nuremberg, Erlangen, Bavaria, Germany
[2] Department of Communication Sciences and Disorders, Louisiana State University, Baton Rouge, Louisiana, USA
[3] Division of Phoniatrics and Pediatric Audiology, Department of Otorhinolaryngology, Head and Neck Surgery, University Hospital Munich, Ludwig-Maximilian-University, Munich, Bavaria, Germany

***Keywords*: Voice Disorder; High-speed Videoendoscopy; Machine Learning**

Abstract:

Objectives / Introduction:
Functional dysphonia (FD) refers to an impairment of voice production, characterized by limitations in vocal performance and acute or persistent changes in voice quality. Due to its diverse genesis in the absence of primary organic findings, there is currently no consensus on the visual assessment of FD. The use of quantitative methods could aid clinicians in standardizing the diagnosis of FD. High-speed videoendoscopy (HSV) is a promising method for the objective evaluation of voice disorders, as its high resolution (e.g. ≥4000 frames per second) allows the detailed analysis of vocal fold vibrations. In this study, we propose a machine learning based approach to objectively assess voice quality using parameters calculated from high-speed endoscopic videos. Our primary focus is to investigate the relationship between the vibratory characteristics of the vocal folds and the resulting voice quality.

Methods:
We gathered HSV recordings of the sustained vowel /i/ from both healthy subjects and patients with functional dysphonia. All recordings have an assigned hoarseness rating $H \in [0, 1, 2, 3]$, which was determined subjectively by an expert based on continuous speech of the respective subject. Glottis segmentation is used to determine the glottal area waveform (GAW) for 250ms of each recording. Subsequently, voice parameters describing glottal dynamics, mechanics and symmetry as well as signal periodicity and harmonicity were computed. We employed machine learning to classify HSV recordings into two levels of hoarseness $H < 2$ (normal, mild) and $H \geq 2$ (moderate, severe), using the resulting output probabilities as interval-scaled severity ratings $\hat{y} \in [0, 1]$. Additionally, glottal features significant for classification were identified using feature selection methods.

Results/Conclusions:
The resulting classification model was evaluated regarding classification performance as well as correlation between predicted output probabilities and subjectively determined hoarseness ratings. In addition, relevant features were analyzed in terms of their correlation with hoarseness.

Acknowledgements:

This research was supported by Deutsche Forschungsgemeinschaft Grant Nos. DO1247/8-2 and SCHU3441/3-2.

THE USE OF 3D HIGH-SPEED VIDEOENDOSCOPY FOR EVALUATING VOCAL FOLD VIBRATORY CHARACTERISTICS AS A FUNCTION OF FREQUENCY

Rita R. Patel[1], Michael Döllinger[2], Bernhard Jakubaß[2], Marion Semmler[2]

[1] Department of Otolaryngology Head and Neck Surgery, Indiana University, Indianapolis, Bloomington, USA
[2] Division of Phoniatrics and Pediatric Audiology at the Department of Otorhinolaryngology Head & Neck Surgery, University Hospital Erlangen, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

***Keywords*: e.g. High-speed videoendoscopy; 3D Imaging, Laser Endoscopy; Vocal fold vibrations**

Abstract:

Objectives / Introduction: Commonly used clinical imaging modalities of videostroboscopy and high-speed videoendoscopy capture vocal fold vibratory motion in only 2D. Several, attempts have been made clinically to capture the 3D vibratory motion clinically with the use of custom laser coupled with either laryngeal imaging systems of high-speed videoendoscopy[1,2,3] and videokymography.[4,5] In 10 vocally healthy adults, Semmler et al (2016)[6] reconstructed the superior surface of the vocal folds using 18 X 18 laser dots[7] coupled with high-sped videoendoscopy. The findings from Semmler et al (2016)[6] in 10 vocally healthy adults revealed that the vertical parameters of amplitude, velocity maximum opening, and velocity maximum closing were at least 50% larger compared to their lateral counterparts during modal phonation at the mid-membranous section of the vocal folds. The current study extends this literature by evaluating 3D vertical vibratory motion at low, high, and modal vocal frequency in large number of participants. The goal of this study was to examine 3D vertical vibratory motion of the vocal folds as a function of changes in vocal frequency.

Methods: A total of 23 vocally healthy individuals (12 males; Mean = 24.8 ± 4.11 years and 11 females; Mean = 23.83 ± 2.76 years) were recruited at Indiana University. High-speed videoendoscopy (PENTAX Medical Model 9710, Montvale, New Jersey) along with custom developed green light laser (Bayerisches Laserzentrum GmbH, Erlangen, Germany) with 18 x 18 dots[7] was used to capture vocal fold dynamics during sustained phonation of the vowel /i:/ at low, middle, and high vocal frequency. The temporal resolution of the high-speed videoendoscopy was 4000 fps and the spatial resolution was 512 x 256 pixels. The low, middle, and high vocal frequency levels were obtained from participants 10%, 20%, and 80% of the pitch range respectively. The 3D vibratory motion was examined along the anterior, middle, and posterior sections of the vocal folds. Simultaneous acoustic signal was captured at 50kHz with a fixed mouth-to-microphone distance of 5inches.

A total of 500 frames were initial segmented using the Glottis Analysis Tools (v. 2018, Erlangen, Germany)[8,9] to obtain the 2D glottal area waveform. Subsequently, a semi-automated custom software called VideoClick (v.2. Erlangen, Germany) was used to detect the 2D pixel position of all laser dots in the initial frame. The 2D laser points were then tracked semi automatically or manually for the entire 500 frames. Each file required approximately 6 to 7 hours for analysis. The 3D coordinates of every laser point were reconstructed from the 2D position using a homography calibration matrix. The 3D parameters of amplitude (mm), maximum velocity (mm/sec), and mean velocity (mm/sec) were computed for the lateral and vertical vibratory motion along the anterior, middle, and posterior sections of the vocal folds for 10%, 20%, and 80% of the vocal frequency ranges.

Statistical analyses were conducted using Linear mixed model method, to evaluate the differences in (a) vocal frequency levels (high vs. normal vs. low pitch), (b) axis level (vertical vs. lateral), (c) position level (anterior vs. middle vs. posterior), and (d) gender differences (male vs. female), with participant as random effect. Pairwise comparisons were adjusted using the Tukey test. P-values were two-tailed and considered significant at values <0.05. All analyses were performed using R, version 4.2.2.

Results: The mean vocal frequency (Hz) across the three tasks of high pitch, normal pitch, and low pitch were statistically significant different $p<.01$; indicating accurate task performance.

Overall, the vertical motion of the superior surface of the vocal fold was consistently larger compared to the lateral motion. Changes in pitch resulted in systematic variations in the maximum amplitude (mm), maximum opening velocity, and maximum closing velocity (mm/sec) in the vertical direction. These findings indicate that the amplitude, maximum opening and closing velocity are large for low pitch and smallest for high pitch in the vertical direction. The position (anterior, middle, or posterior) on the superior surface of the vocal fold matters exclusively during vocal fold closing and not during opening. Along the superior surface, the mean and maximum closing velocities are greater posteriorly for low pitch compared to middle and anterior, but not for normal or high pitch.

Conclusions: The study highlights the importance of measuring the vertical motion of the superior surface of the vocal fold. Along the superior surface, the mean and maximum closing velocities were greater posteriorly for low pitch.

Acknowledgements:

The authors acknowledge contributions from Caroline Riefe and Tzu Pei Tsai at Indiana University for their assistance with data segmentation and analyses. Contributions by R Patel were supported by NIH NIDCD R01DC017923 and by M Semmler & M Döllinger were supported by Deutsche Forschungsgemeinschaft (DFG) under grant no. DO1247/16-1.

References:

1. Patel, R., Donohue KD, Lau D, Unnikrishnan H. In vivo measurement of pediatric vocal fold motion using structured light laser projection. *Journal of Voice.* 2013;27(4):463-472.
2. Patel, R., Donohue KD, Johnson WC, Archer SM. Laser projection imaging for measurement of pediatric voice. *The Laryngoscope.* Nov 2011;121(11):2411-2417.
3. de Mul FF, George NA, Qiu Q, Rakhorst G, Schutte HK. Depth-kymography of vocal fold vibrations: part II. Simulations and direct comparisons with 3D profile measurements. *Phys Med Biol.* Jul 7 2009;54(13):3955-3977.
4. George NA, de Mul FF, Qiu Q, Rakhorst G, Schutte HK. New laryngoscope for quantitative high-speed imaging of human vocal folds vibration in the horizontal and vertical direction. *J Biomed Opt.* Nov-Dec 2008;13(6):064024.
5. Semmler M, Kniesburges S, Birk V, Ziethe A, Patel R, Döllinger M. 3D Reconstruction of Human Laryngeal Dynamics Based on Endoscopic High-Speed Recordings. *IEEE Trans Med Imaging.* Jul 2016;35(7):1615-1624.
6. Semmler M, Döllinger M, Patel RR, Ziethe A, Schutzenberger A. Clinical relevance of endoscopic three-dimensional imaging for quantitative assessment of phonation. *Laryngoscope.* Mar 14 2018;128:2367-2374.
7. Lohscheller J, Toy H, Rosanowski F, Eysholdt U, Döllinger M. Clinically evaluated procedure for the reconstruction of vocal fold vibrations from endoscopic digital high-speed videos. *Medical image analysis.* Aug 2007;11(4):400-413.
8. 28. Kist AM, Gomez P, Dubrovskiy D, et al. A Deep Learning Enhanced Novel Software Tool for Laryngeal Dynamics Analysis. *J Speech Lang Hear Res.* Jun 4 2021;64(6):1889-1903.

RELIABILITY OF SEGMENTATION OF VOCAL FOLD EDGES FROM HIGH-SPEED VIDEOENDOSCOPY AND THE PARAMETERS AFFECTING IT

Hamzeh Ghasemzadeh[1,2,3], Bernhard Jakubaß[3], Maria E. Powell[4], David S. Ford[5], Dimitar D. Deliyski[3]

[1] Center for Laryngeal Surgery and Voice Rehabilitation, Massachusetts General Hospital, Boston, MA, USA
[2] Department of Surgery, Harvard Medical School, Boston, MA, USA
[3] Department of Communicative Sciences and Disorders, Michigan State University, East Lansing, MI, USA
[4] Department of Otolaryngology-Head & Neck Surgery, Vanderbilt University Medical Center, Nashville, TN, USA
[5] Department of Speech-Language Pathology, Duquesne University, Pittsburgh, PA, USA

***Keywords*: Instrumental Assessment of Voice; Spatial Segmentation; Ground Truth; Reliability Analysis**

Objectives / Introduction:
Biomechanical properties of the vocal folds (e.g., stiffness) and biological changes during their development can indirectly be estimated from the velocity of the vocal folds. High-speed videoendoscopy (HSV) can provide images with high spatial and high temporal resolution and, hence, is an ideal modality for computing velocity measures from vocal folds. Appropriate validation and evaluation of such velocity measures require the existence of reliable ground truths, which in turn depends on the existence of reliable segmentation of the vocal fold edges. However, the parameters affecting the reliability and validity of segmentation of vocal fold edges have not been studied very well. This study presents a framework for creating reliable ground truth for segmentation of vocal fold edges. It also quantifies the effect of different parameters on reliability of such ground truth. The findings of this study can also provide guidance for designing more accurate and robust methods for automated segmentation of vocal fold edges.

Methods:
A three-stage framework is proposed to create reliable ground truth for segmentation of vocal fold edges. In the first stage, three laryngeal imaging experts performed the segmentation task independently. In the second stage, regions with high discrepancies between experts were identified. The regions with high discrepancy from each expert were randomly divided between the other two experts, and they were tasked to make proper adjustments and modifications to the initial segmentation within disparity regions. In the third stage, the remaining regions with high discrepancies were adjusted by consensus between the three experts. One hundred consecutive frames from 12 HSV recordings were segmented based on this framework. The samples were equally balanced between males and females and also between vocally typical individuals and patients with voice disorders. Ten frames from each video were randomly re-segmented to allow for evaluation of segmentation reliability at different stages of the framework. The analyses were based on edge variability which was defined as the maximum lateral difference between the 3 segmented vocal fold edges. This metric is inversely related to segmentation reliability and directly related to segmentation uncertainty. In the first experiment, inter and intra-rater edge variability at different stages of the framework were evaluated. In the second experiment, the effects of different glottal phases (i.e., temporal dependency) and different sections of vocal folds along the anterior-posterior direction (i.e., spatial dependency) on uncertainty of segmentation were studied. In the third experiment, the effects of sex (male vs female) and diagnosis (patients vs controls) on uncertainty of segmentation were studied.

Results:
Experiment 1: The intra-rater edge variability at the first stage was 0.30±0.17, which improved to 0.25±0.12 after the reconciliation step. The inter-rater edge variability at the first stage was 0.58±0.29, which improved to 0.41±0.19 after the reconciliation step.
Experiment 2: Inter-rater edge variability was non-uniform along the length of the glottis, with significantly higher uncertainty in the anterior and posterior sections of the glottis. Additionally, edge variability was inversely correlated with glottal opening, meaning vocal fold edges from open phases of the glottis had higher segmentation uncertainties.
Experiment 3: Diagnosis had a significant effect, where vocal fold edges from patients had significantly higher uncertainty compared to vocally typical controls. The effect of sex on the uncertainty of segmentation of vocal fold edges was non-significant.

Conclusions:
The proposed framework generated highly reliable ground truth for evaluating the validity of automated spatial segmentation methods. The uncertainty of segmentation of the vocal fold edges depended on the glottal phase, the location of the edge along the anterior-posterior direction, and the diagnosis. These factors need to be considered during the evaluation of spatial segmentation methods. They also can be used to design more reliable automated segmentation methods.

Acknowledgements:
This research was funded by the National Institutes of Health, National Institute on Deafness and Other Communication Disorders (R01DC017923, R01DC019402).

COMPARING TEN VOCAL FOLD CLOSURES USING STROBOSCOPY, HIGH-SPEED VIDEOGRAPHY, EGG, AND FREQUENCY ANALYSIS: A WORK IN PROGRESS

Lisa Popeil[1], Zhaoyan Zhang[2], Dinesh Chhetri[2], Sarah Lehoux[2], James P. Thomas[3]

[1]Voiceworks, Los Angeles, California, USA

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

[3]MD ENT - Portland Oregon, USA

Keywords: Voice; Singing, Vocal Fold Adduction, Vocal Pedagogy

Abstract:

Objectives / Introduction: The goal of this project is to explore if singers and speakers can consistently produce more precise differentiations vocally than just the three typical phonation modes, i.e. breathy, flow and pressed. The first author has attempted to refine the terminology for pedagogical purposes, culminating in a model of vocal fold adduction which describes ten unique vocal fold closure types. These terms compare “squeezed”; 4 types of “hard" (front, mid, back and even); “clean”; “blowy”; and 3 types of “breathy" (slightly breathy, moderately breathy and very breathy). The purpose of this ongoing research project is to analyze and compare the acoustic and production characteristics between these ten closure types in one professional singer and voice pedagogue[1]

Methods: Acoustics, electroglottography (EGG), and high-speed videoendoscopy were recorded in which a single subject produced samples of the ten vocal fold closure types. The audio recordings will then be used in a perceptual study to verify that the ten vocal cord closures can be consistently identified by expert listeners. The closed and speed quotients of these vocal fold closures will be extracted from the EGG data.

Results: In this presentation, stroboscopy and high-speed video, audio samples, and EGG findings will be shared, with a focus on co-variations between glottal and supraglottal adjustments between the ten closure types.

Conclusions: This model of vocal fold closures has proven eminently teachable not only in the artistic singing realm but also for the training of speech pathologists in working with hyper and hypo-functional speaking voices. It’s important to precisely identify, and correct, adduction problems based on sensation and sound resulting in heightened vocal fold control in singing, as well as guiding a voice-user how to avoid harmful vocal fold pressing.

MALE AND FEMALE VOICES SHOW DIFFERENCES IN SPECTRAL DEVELOPMENT OVER FOUR YEARS OF CONSERVATORY TRAINING

Reuben Scott Walker[1,2] , Mario Fleischer[1], Johan Sundberg[3], and Dirk Mürbe[1,2]

[1] Department of Audiology and Phoniatrics, Charité – Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany

[2] Hochschule für Musik Carl Maria von Weber Dresden, Dresden, Germany

[3] Department of Speech, Music and Hearing, School of Computer Science and Communication, Royal Institute of Technology, Stockholm, Sweden

***Keywords*: Voice; Acoustics; Longitudinal**

Abstract:

Objectives / Introduction:

Male and female classical vocal timbre differs due to a variety of reasons, including the frequency range of sung repertoire and physiological differences [1]. This paper sets out to use spectral measures to investigate the question: Do classical singers develop similarly over the course of their conservatory training or does their spectral development exhibit differences based on grouping them in soprano/mezzo/contralto or tenor/baritone/bass voice classifications?

Methods:

A series of vocal exercises were recorded yearly between 2002 and 2021 by each undergraduate student at the Hochschule für Musik Carl Maria von Weber Dresden. From the 227 available classical voice students, 187 had had recordings in multiple years of study and were included in the analysis (F=114, M=73). For each student, we selected two of the yearly recorded exercises for retrospective longitudinal analysis. First we extracted the middle 50% of the sustained fifth from an ascending/descending triad exercise performed on an /a/ vowel. For a second analysis, we used the entirety of “Avezzo a vivere”, a one octave exercise from Nicolo Vaccai’s *Metodo pratico di canto.* Two measures were calculated for each sample: alpha ratio, the proportion of acoustic energy above 1000 Hz to acoustic energy below 1000 Hz, and H1H2LTAS, a ratio of averaged levels from two frequency bands of the long-term average spectrum (LTAS); one band for the octave of the sung exercise and the second band one octave above [2]. To include countertenor voices (n=7), the samples were split into binary groups by the frequency level of their sung exercises, either soprano/mezzo/alto or tenor/baritone/bass.

Results:

There was a significant interaction for voice classification and time for H1H2LTAS for both sustained phonation and the repertoire sample. There was a significant interaction for alpha ratio for sustained phonation. There was no interaction for alpha ratio in the repertoire sample. Female and countertenor H1H2LTAS increased for both sustained phonation and the repertoire sample. Female and countertenor alpha ratio did not change for sustained phonation but did increase for the repertoire sample. Male H1H2LTAS did not change for sustained phonation or the repertoire sample. Male alpha ratio increased for both sustained phonation and the repertoire sample.

Conclusions:

Spectral development differed between female/countertenor and male students during conservatory training. Female and countertenor students exhibited increases in ratios of low frequency to high frequency energy for the sustained vowel, suggesting a relative strengthening of the fundamental frequency during vocal training, while male voices exhibited the inverse relationship, suggesting that their development is characterized by a relative strengthening of higher frequency partials.

Acknowledgements:

We thank Hartmut Zabel for input regarding the maintenance of the acoustic database.

References:

[1] Titze, I. R. (1989). Physiologic and acoustic differences between male and female voices. The Journal of the Acoustical Society of America, 85(4), 1699-1707.

[2] Scherer, K. R., Sundberg, J., Tamarit, L., & Salomão, G. L. (2015). Comparing the acoustic expression of emotion in the speaking and the singing voice. *Computer Speech & Language*, 29(1), 218-235.

Articulatory and acoustic differences between lyric and dramatic singing in western classical music

Matthias Echternach [1], Fabian Burk [2], Jonas Kirsch[1], Louisa Traser [3], Peter Birkholz [4], Michael Burdumy [5], Bernhard Richter [3]

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

[2] Department of Otorhinolaryngology and Plastic Surgery, SRH Wald-Klinikum Gera, Gera, Germany

[3] Institute of Musicians' Medicine, Faculty of Medicine, Freiburg University and Freiburg University Medical Center, Freiburg, Germany

[4] Institute of Acoustics and Speech Communication,Technische Universität Dresden, Dresden, Germany

[5] Department of Medical Physics, Radiology, and Institute of Musicians' Medicine, Faculty of Medicine, Freiburg University and Freiburg University Medical Center, Freiburg, Germany

***Keywords*: Singing – Vocal tract –Vocal Category – Singer`s Fach**

Abstract:

Introduction: Within the realm of voice classification, singers could be sub-categorized by the weight of their repertoire, the so called "singers` Fach". However, the opposite pole terms "lyric" and "dramatic" singing are not yet well defined.

Material and Methods: 9 professional singers of different singers' Fach were asked to sing a diatonic scale on the vowel /a/, first in what the singers considered as lyric, and second considered as dramatic. Image recording was performed using real time MRI with 25 frames/s and the audio signal was recorded via an optical microphone system. Analysis was performed with regard to sound pressure level (SPL), vibrato amplitude and frequency and resonance frequencies as well as articulatory settings of the vocal tract.

Results: The analysis revealed three primary differences between dramatic and lyric singing: Dramatic singing was associated with greater SPL, greater vibrato amplitude and frequency as well as lower resonance frequencies. However, all these strategies showed a considerable individual variability

Conclusion: The singers` Fach might contribute to perceptual differences even for the same singer with regard to the respective repertoire.

Acknowledgements:

The work was supported by Deutsche Forschungsgemeinschaft, grants Ec409/1-4

PRECISION IN ARTICULATORS AND PHONATION WHILE PERFORMING STACCATO SCALES AT DIFFERENT SPEEDS COMPARED TO LEGATO SINGING

Marie Köberlein [1], Jonas Kirsch[1], Fabian Burk[2], Louisa Traser[3], Bernhard Richter[3], Michael Burdumy[4], Matthias Echternach [1]

[1]Department of Otorhinolaryngology, Division of Phoniatrics and Pediatric Audiology, Munich University Hospital (LMU Munich), Munich, Germany
[2]Department of Otorhinolaryngology and Plastic Surgery, SRH Wald-Klinikum Gera, Gera, Germany
[3] Institute of Musicians' Medicine, Medical Center, University of Freiburg, Freiburg, Germany
[4]Department of Medical Physics, Radiology, University of Freiburg, Freiburg, Germany

***Keywords*: Articulation; Vocal tract; Magnetic resonance imaging**

Abstract:

Objectives / Introduction: During phonation, the vocal tract configuration determines the resonance settings, i.e., vowel qualities, timbre and loudness[1-3], and also plays a role in interactions with the voice source[4,5]. Staccato exercises, characterized as phonation with separated tones with a pause in between, are frequently used in voice training in order to improve precision in fundamental frequency (f_o) and articulator coordination[6]. However, the impact of speed in staccato on vocal tract adjustments and phonation precision, had not been investigated, yet. It was hypothesized that the behavior when singing separated tones would generally differ from singing uninterrupted, connected tones (legato) and that the behavior of staccato would differ between different speed conditions.

Methods: A total of 12 professional singers, comprising 4 sopranos, 3 mezzo-sopranos, 3 tenors, and 2 baritones/basses, underwent real-time 2D-MRI recording at 25 frames per second. They sang a scale of 6 ascending diatonic pitches on vowel [a:] in legato and various staccato speeds (60, 120, 180, and 240 beats per minute). Measurements of lip opening (LO), jaw opening (JO), jaw protrusion (JP), tongue position (HPT), pharynx width (PW), and larynx position (LP) were obtained from the MRI data as described earlier[7]. Additionally, fundamental frequency analysis was performed on the concurrently recorded audio signal after noise cancellation, considering the peak 6 dB bandwidth as a measure of pitch stationarity.

Results: The articulator measurements reveal minimal differences in absolute values between legato and staccato phonation, even though being statistically significant for some articulators (p-values: JO: 0.049; HPT: 0.013; PW: 0.047, LP: 0.022. Median differences in mm: JO: 1.19 HPT: 1.75; PW: 0.21; LP: 1.03). No great variations were observed between the fastest and slowest staccato phonation, despite JO being statistically significant (JO: p=0.001, Median difference: 0.70 mm). Noteworthy vocal tract adjustments were only discernible during pauses between staccato notes and pauses in the slowest staccato task but with median differences staying below 2.17 mm for all articulators (Mean Median difference of all articulators: 1 mm). The f_o analysis showed an increasing imprecision from legato (least imprecision with average pitch deviation of 9.4 cents) to fast staccato (highest imprecision with average pitch deviation of 53.4 cents).

Conclusions: Comparing staccato and legato phonation in professional operatic singers, there are only negligible differences in vocal tract shapes, and the variation in staccato speeds does not lead to significant alterations in vocal tract configurations. A minor pause behavior with small changes, interpreted as relaxation, occurs only in slow staccato tasks. The f_o precision decreases with accelerating tasks, which is rather linked to laryngeal mechanisms than to articulation.

Acknowledgements:

Matthias Echternach (grant Ec409/1-4 and Ec409/5-1) and Bernhard Richter (grant Ri 1050/4-1) are supported by the Deutsche Forschungsgemeinschaft (DFG).

References:

1. Stevens KN. Acoustic Phonetics. Cambridge: MIT Press, 1998.
2. Sundberg J. The Science of the Singing Voice. Northern Illinois University Press, 1987.
3. Titze IR. Principles of voice production. Prentice Hall, NJ, 1994.
4. Maxfield L, Palaparthi A, Titze I. New Evidence That Nonlinear Source-Filter Coupling Affects Harmonic Intensity and fo Stability During Instances of Harmonics Crossing Formants. J Voice 2016.
5. Zanartu M, Mehta DD, Ho JC, Wodicka GR, Hillman RE. Observation and analysis of in vivo vocal fold tissue instabilities produced by nonlinear source-filter coupling: a case study. J Acoust Soc Am 2011; 129:326-339.
6. Miller R. Solutions for Singers: Tools for Performers and Teachers. Oxford University Press, 2004, p.8.
7. Echternach M, Burk F, Burdumy M, Traser L, Richter B. Morphometric Differences of Vocal Tract Articulators in Different Loudness Conditions in Singing. PloS one 2016; 11:e0153792.

EXAMINATION OF DIAPHRAGM CONTROL AND ITS EFFECTS ON PITCH LEAP DURING OPERA SINGING USING REAL-TIME MRI

Natsuki Toda[1], Hironori Takemoto[1], Jun Takahashi[2], Seiji Adachi[3]

[1] Chiba Institute of Technology, Chiba, Japan
[2] Osaka University of Arts, Osaka, Japan
[3] Tianjin University, Tianjin, China

***Keywords*: Opera singing, Real-time MRI, Diaphragm, Subglottal Pressure**

Abstract:

Objectives / Introduction: Opera singers use techniques for controlling vocal tract resonance, such as the singer's formant [1] and formant tuning [2], to produce a *squillo* or ringing voice. In addition, they are able to control the subglottal pressure to change the intensity of phonation. Vocal tract resonance is determined by the shape of the vocal tract while subglottal pressure is determined by changes in lung volume. However, it is difficult to observe the shape of the vocal tract and changes in lung volume directly.

Recently, real-time magnetic resonance imaging (rtMRI) was used to record articulatory movements during speech [3]. This technology can also be applied to lungs, as we demonstrated in a pilot study. In the previous work, rtMRI was used to record the changes in the right lung shape in the sagittal plane of two professional opera singers as they were singing. Then, the changes in the lung shape and diaphragm position were extracted and analyzed from the rtMRI videos [4]. We observed that, although the singers gradually raise the diaphragm to decrease the lung volume while singing, they temporarily lower the dorsal side of the diaphragm to increase the lung volume when there is a drop in pitch [4]. After a temporary lowering, the singers again raised the diaphragm. Through this process, no interruptions in the singing voice were perceived. In the present study, we examined why singers lower their diaphragm during singing by estimating the subglottal pressure.

Methods: The participants were two professional opera singers, a tenor, and a baritone. Fig. 1 shows a singing task consisting of three perfect 5th leaps (①-③) and four octave leaps (④-⑦). The participants performed the singing task in the supine position while undergoing an MRI scan (Siemens MAGNETOM Prisma fit 3T installed at ATR Promotions). While singing, movements of the right lung in the sagittal plane were recorded as a video at a speed of 10 fps for approximately 50 s. The pixel resolution was 1.22 x 1.22 mm and the slice thickness was 10 mm. The singing voice during scanning was recorded using an optical microphone (Optoacoustics, Optimic1140). In addition, because the sound recorded includes the loud scanning noise from the MRI, the singing voice for the same task was recorded in the supine position in a quiet room.

From each video frame, the contour of the right lung was semiautomatically extracted using the machine learning library Dlib [5], as shown in Fig. 2. The area of the right lung was calculated using the contours. Subglottal pressure was estimated from changes in lung area. First, we assumed that the lung volume was proportional to the 3/2 power of the lung area, and the maximum volume was 5,000 cm^3. Based on these assumptions, changes in lung volume could be estimated at 0.1 s intervals. During frame switching from inhalation to phonation, the lung pressure can be considered equal to the atmospheric pressure. At the beginning of the next frame, we assume that the lung volume suddenly changes but remains constant in the frame and that the subglottal pressure changes accordingly by adiabatic compression. In this frame, the mass of airflow per second can be calculated using the following equation:

$$m = \frac{Ap}{\sqrt{RT}} \sqrt{\frac{2\gamma}{\gamma - 1} \cdot \left(\frac{p_!}{p}\right)^{\frac{2}{\gamma}} - \left(\frac{p_!}{p}\right)^{\frac{\gamma+1}{\gamma}}} \quad \text{kg/s},$$

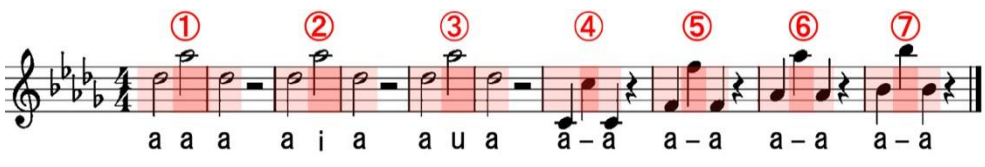


Fig. 1 Singing task

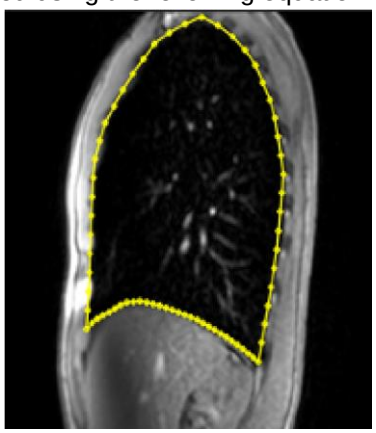


Fig. 2 Lung contour

where A is the time averaged glottal area, 0.05 cm^2, p is the subglottal pressure which is constant until the end of the frame, R is the gas constant of air, 287.1 J/(kgK), T is the temperature of air, 310 K, γ is the specific heat ratio of air, 1.4, and p_0 is atmospheric pressure,100 kPa. At the end of the frame, the subglottal pressure changes by $-\dot{m}\,\Delta t RT/V$, where Δt is the frame interval, 0.1 s, and V is the lung volume in the frame. Assuming that the above process is repeated during phonation, changes in subglottal pressure can be estimated.

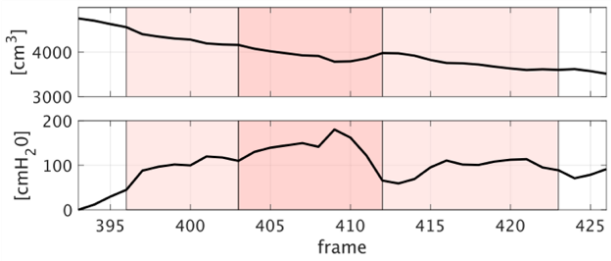


Fig. 3 Estimated lung volume (upper) and subglottal pressure (lower) for tenor

Results: Fig. 3 shows the estimated lung volume and subglottal pressure (gauge pressure) of the tenor during the octave leap at ⑥. Although the lung volume tended to decrease overall, it slightly increased from 410th frame to 412th frame, just before the pitch leaped down. The subglottal pressure was high at the high pitch (dark shaded), while it was low at the low pitch (light shaded). This result matches the fact that a high pitch requires high pressure, and a low pitch requires low pressure [6]. If air compressibility is not considered, the increase in lung volume from the 410th frame to 412th frame would cause negative pressure, that is, inhalation. However, in the present study, because air compressibility was considered, the subglottal pressure rapidly decreased, but was still positive for continuing singing. A similar pressure decrease was found for the baritone during the octave leap at ⑥ (Fig. 4). Based on these results, it is confirmed that professional singers can rapidly decrease the subglottal pressure to an extent suitable for a low pitch by intentionally lowering the diaphragm, even during singing. This prediction is supported by the waveform of a singing voice recorded in a quiet room. Fig. 5 shows the waveforms of the two singers. After the pitch leaped down, the amplitude decreased. This matches the estimated subglottal pressure pattern.

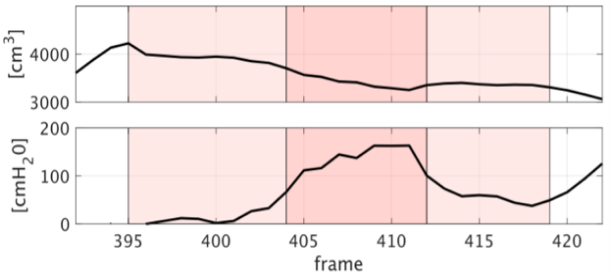


Fig. 4 Estimated lung volume (upper) and subglottal pressure (lower) for baritone

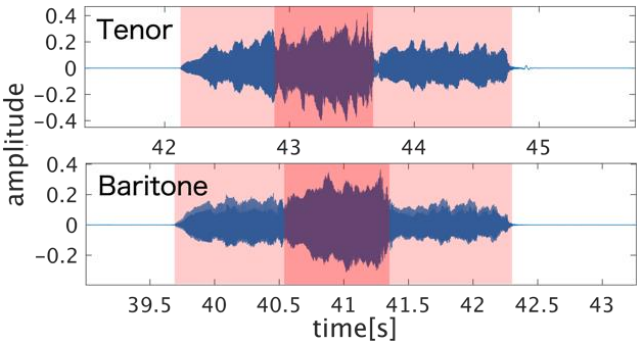


Fig. 5 Sound waveforms

Conclusions: It has long been assumed that singers continue to raise their diaphragm to generate an expiratory flow during singing. However, we demonstrated previously that singers temporarily lowered the diaphragm even while singing [4]. The reason why singers adopted this technique was examined in this study. From the rtMRI data, changes in subglottal pressure were estimated by considering air compressibility. The results indicated that professional opera singers can rapidly decrease the high subglottal pressure suitable for the high pitch to a low pressure during the leap to the low pitch by lowering the diaphragm to increase the lung volume. According to personal interviews, some singers refer to this technique as “re-supporting.” “Breath support” might be what singers feel when raising the diaphragm appropriately for pitch and loudness. Therefore, “re-support” could be what singers feel when raising the diaphragm again after temporarily lowering.

Acknowledgements:

This research was partly supported by JSPS KAKENHI (grant number 23K11172).

References:

[1] J. Sundberg, *THE SCIENCE OF THE SINGING VOICE*, Northern Illinois University Press, 1987.

[2] Garnier *et al.*, “Vocal tract adjustments in the high soprano range,” The Journal of the Acoustical Society of America, Volume127, Number6, pp. 3771–3780, 2010.

[3] Takemoto *et al*., “Speech Organ Contour Extraction using Real-Time MRI and Machine Learning Method,” Proc. Interspeech 2019, pp.904-908, Sep.2019.

[4] Toda *et al*., “Examination of exhalation control in opera singing using real-time MRI,” The Journal of the Acoustical Society of Japan, Volume78, Number11, pp. 646–649, 2022, (in Japanese).

[5] D. E. King. “Dlib-ml: A Machine Learning Toolkit,” Journal of Machine Learning Research 10, pp. 1755–1758, 2009.

[6] Cleveland. T, J.Sundberg, “Acoustic analysis of three male voices of different quality,” In Proc. of Stockholm Music Acoustics 1983(SMAC83), no.1, pp.143-156, Jul.1983.

ACOUSTIC CONTRIBUTIONS OF THE HYPOPHARYNGEAL CAVITIES IN GENERATING SINGER'S FORMANT: A CASE STUDY

Hironori Takemoto[1], Itsuki Shishime[1], Natsuki Toda[1], Jun Takahashi[2], Yukiko Nota[3]

[1] Chiba Institute of Technology, Chiba, Japan
[2] Osaka University of Arts, Osaka, Japan
[3] National Institute of Japanese Language and Linguistics, Tokyo, Japan

***Keywords*: Singer's Formants, Laryngeal Cavity, Piriform Fossae, FDTD**

Abstract:

Objectives / Introduction: The singer's formant, a spectral emphasis around 3 kHz commonly present in vowels produced by male opera singers [1], has been theorized to result from a cluster of third, fourth, and fifth formants (F3, F4, and F5) [1]. The hypopharyngeal cavities, encompassing the laryngeal cavity and bilateral piriform fossae, generates this cluster [1]. However, the specific mechanisms remain ambiguous. This case study investigates the acoustic contributions of the laryngeal cavity and piriform fossae to the generation of the singer's formant. We employed MRI (Magnetic Resonance Imaging) to capture the three-dimensional vocal tract shape of a tenor during singing and subsequently analyzed transfer functions when modifying the hypopharyngeal cavities.

Methods: In an MRI scanner (Siemens, MAGNETOM Prisma fit 3T), a professional tenor vocalized the vowel /a/ while supine for 15 s at a pitch corresponding to F3 (174.6 Hz). Scanning commenced approximately 2 s after initiating singing. The singing voice was recorded using an optical microphone (Optoacoustics Optimic 1140). The vowel spectrum was extracted from recordings obtained before scanning due to loud noise during scanning. Volumetric data of the vocal tract were reconstructed from the images, with a pixel resolution and slice thickness of 1 × 1 mm and 2.5 mm, respectively. Dental images were superimposed onto the volume data [2], and the vocal tract shape from the glottis to the lips was extracted. The vocal tract wall thickness was increased outward by 3 mm to maintain the vocal tract shape, resulting in a vocal tract model [3]. The model was discretized to enhance spatial resolution to 0.5 mm. The transfer function from the glottis to the lips in the original model was calculated using the finite-difference time-domain method [3]. Subsequently, the laryngeal cavity was excluded from the original model, and the transfer function from the pharynx to the lips was computed. The bilateral piriform fossae were omitted from the original model, and the transfer function from the glottis to the lips was calculated. Henceforth, these transfer functions are denoted as TF-org, TF-lc, and TF-pf (Fig. 1).

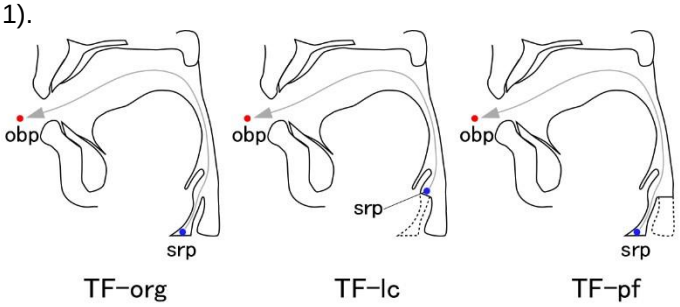


Fig. 1 Calculation of the three transfer functions (srp: source point, obp: observation point)

Results: Fig. 2 displays the TF-org alongside the spectrum of the sung vowel. TF-org aligned closely with the spectrum below 3.5 kHz. The singer's formant, representing a formant cluster of F3, F4, and F5, was evident on the spectrum in the frequency range from 2.3 kHz to 3.1 kHz. Correspondingly, TF-org exhibited clusters of the third, fourth, and fifth resonances (fR3, fR4, and fR5, respectively) within the same frequency range. Hence, modifications to the original vocal tract model offered insights into the mechanism underlying the generation of the singer's formant.

Fig. 3 presents the transfer functions: TF-org, TF-lc, and TF-pf. A comparison between TF-org and TF-lc revealed the role of the laryngeal cavity in generating the singer's formant. Removing the laryngeal cavity induced significant changes in the resonance peak cluster: fR4 disappeared, fR3 underwent a frequency increase of 50 Hz and a level decrease of 23 dB, and fR5 decreased in frequency and level by 100 Hz and 29 dB, respectively. However, this removal resulted in only minor changes in the other peaks and dips. These findings suggest that the laryngeal cavity generates fR4—potentially the principal component of the singer's formant—with limited impact on other resonances and anti-resonances, as corroborated by a previous study [4]. This phenomenon arises from the upper part of the

laryngeal cavity constricting and the lower part of the pharyngeal cavity expanding during opera singing, rendering the laryngeal cavity acoustically independent from the remainder of the vocal tract.

A comparison between TF-org and TF-pf showed the contribution of piriform fossae to generating the singer's formant. The elimination of the piriform fossae removed a deep spectral dip at 3.5 kHz and impacted the resonance peak cluster as follows: fR3 increased in frequency by 50 Hz and in level by 4 dB, fR4 increased only in level by 1 dB, and fR5 increased in frequency by 250 Hz and in level by 4 dB. These findings indicate that the vocal tract without piriform fossae generated fR3 and fR5, while the piriform fossae decreased their frequencies. The degree of frequency decrease occurred at fR5, which was near the dip caused by the piriform fossae. The piriform fossae acted as side branches of the vocal tract, generating pole-zero pairs on the transfer function [4]. Consequently, the deep spectral dip at 3.5 kHz on TF-org corresponded to a zero. fR5 plausibly corresponds to the pole, as indicated by Vampola *et al.* [5].

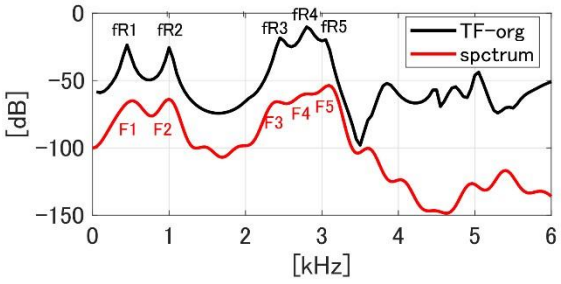


Fig. 2 Original transfer function and vowel spectrum

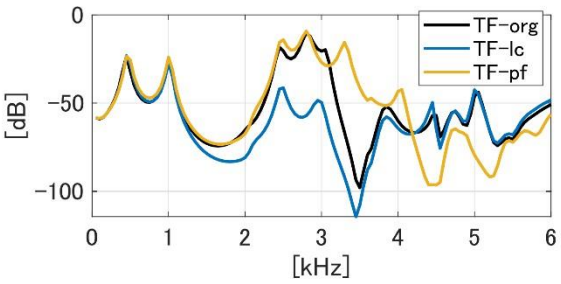


Fig. 3 Three transfer functions

Conclusions: The three-dimensional vocal tract shape of the tenor while singing the vowel /a/ was measured using MRI. The calculated transfer function from the vocal tract shape aligned well with the vowel spectrum below 3.5 kHz, and a resonance peak cluster corresponding to the singer's formant was observed. Transfer functions were computed and analyzed with modification to the hypopharyngeal cavities to explore the acoustic role of the hypopharyngeal cavities in generating the singer's formant. The results indicated that the laryngeal cavity generated fR4, the primary component of the resonance peak cluster, significantly increasing the level of the cluster. However, the piriform fossae generated a deep spectral dip directly above the cluster and decreased the frequency of fR5. Consequently, the clusters became tighter. fR5 may represent a pole of the pole-zero pair generated by the interaction between the piriform fossae and vocal tracts other than the hypopharyngeal cavities.

The mechanism for generating a singer's formant can be hypothesized as follows based on the foregoing discussion: The upper part of the laryngeal cavity narrowed during opera singing, acoustically isolating it from other vocal tract parts. The first resonance of the laryngeal cavity appeared as fR4 on the transfer function. Constricting the upper part of the laryngeal cavity (laryngeal vestibule) and expanding the lower part (laryngeal ventricle) decreased the frequency of fR4 to approximately fR3. The frequency of fR5 can be decreased to a certain degree (e.g., 3.5 kHz) by the deformation of the vocal tract other than the piriform fossae [6]. The piriform fossae elongated owing to lowering the larynx during singing, decreasing dip frequencies from 4 kHz to 3.5 kHz. Hence, the frequency of fR5 can be further decreased to be close to the clusters of fR3 and fR4. Consequently, the resonance peak clusters fR3, fR4, and fR5 were generated around 2.8 kHz.

Acknowledgements:

This work was supported by JSPS KAKENHI, Grant Number 23K11172, and funding from the Center for Language Resource Development, National Institute for Japanese Language and Linguistics.

References:

[1] Sundberg, J. (1974). Articulatory interpretation of the "singing formant." *The Journal of the Acoustical Society of America*, *55*(4), 838–844.

[2] Takemoto, H., Kitamura, T., Nishimoto, H., & Honda, K. (2004). A method of tooth superimposition on MRI data for accurate measurement of vocal tract shape and dimensions. *Acoustical Science and Technology*, *25*(6), 468– 474.

[3] Takemoto, H., Mokhtari, P., & Kitamura, T. (2010). Acoustic analysis of the vocal tract during vowel production by finite-difference time-domain method. *The Journal of the Acoustical Society of America*, *128*(6), 3724–3738.

[4] Takemoto, H., Adachi, S., Kitamura, T., Mokhtari, P., & Honda, K. (2006). Acoustic roles of the laryngeal cavity in vocal tract resonance. *The Journal of the Acoustical Society of America*, *120*(4), 2228–2238.

[5] Vampola, T., Horáček, J., & Švec, J. G. (2015). Modeling the influence of piriform sinuses and valleculae on the vocal tract resonances and antiresonances. *Acta Acustica United with Acustica*, *101*(3), 594–602.

[6] Story, B. H. (2006). Technique for "tuning" vocal tract area functions based on acoustic sensitivity functions. *The Journal of the Acoustical Society of America*, *119*(2), 715–718.

CHARACTERISTICS OF GLOTTAL AND SUPRAGLOTTAL OSCILLATIONS OF TEN DIFFERENT IRREGULAR PHONATION TYPES DURING METAL SINGING

Louisa Traser[1,2*], Mario Fleischer[3*], Daniel Priegnitz[4], Marie Köberlein[5], Jonas Kirsch[5], Johannes Fischer[2,6], Bernhard Richter[1,2], Matthias Echternach[5]

[1]Institute of Musicians' Medicine, Medical Center, University of Freiburg, Freiburg, Germany,
[2]Faculty of Medicine, University of Freiburg, Freiburg, Germany,
[3]Department of Audiology and Phoniatrics, Charité - Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany,
[4]www.vocalsontherocks.de, Freiburg, Germany,
[5]Department of Otorhinolaryngology, Ludwig-Maximilians-Universität München, Division of Phoniatrics and Pediatric Audiology, LMU Klinikum, Munich, Germany,
[6]Department of Radiology, Medical Center, Medical Physics, University of Freiburg, Freiburg, Germany

*These two authors contributed equally to this work

***Keywords*: irregular voice, metal singing, high-speed imaging**

Abstract:

Objectives / Introduction:
Voice production in healthy individuals is characterized by regular vibrations of the vocal folds (VF). Here, supraglottic structures play an important role in acoustics and the interaction with airflow while remaining motionless (1). However, supraglottic vibrations can be found involuntary, e.g. in patients suffering from muscle tension dysphonia (ventricular dysphonia) (2) or voluntary, in patients whose glottal voice source cannot be restored and supraglottic vibrations are trained as substitute (3). Additionally, in pop/rock/metal singing (4–6) and classical Asian singing styles (7,8), a virtuous control of irregular voice qualities with glottal and supraglottal vibration patterns is known. Supraglottic vibrations are characterized by a higher irregularity compared to VF vibrations (9). Literature on irregular singing styles include different descriptions of simultaneous glottal-supraglottal nearly periodic vibrations (7,10), irregular glottal (11), or chaotic supraglottal (12) vibratory mechanisms. Still, principles of these vibration patterns have often been studied in individual singers who have mastered one specific vocal quality. It is therefore uncertain whether the vibratory mechanisms develop on the basis of individual predisposition, e.g. muscle population in the ventricular fold during training, or whether it is possible to achieve conscious control of different irregular vibration strategies of glottis and supraglottis simultaneously. Therefore, we examined the phonatory outcome of different irregular voice qualities in one trained subject.

Methods:
The subject is a professional metal singer and singing teacher (male, age 31) who is trained to produce ten different irregular voice qualities (Vocal Fry, Grunt, Distortion, Undertone, Rattle, Growl, Deathgrowl, Deathshout, Full Distortion, and Fry Scream). The nomenclature used for our study follows that of the singer because no generally accepted nomenclature exists. We used high-speed digital imaging via flexible endoscopy and simultaneous electroglottographic and audio recording for evaluation as described in (13). After transitioning from his normal phonation on vowel [i:], the singer was asked to perform a glissando over the pitch range possible for this quality.
In analogy to the established glottal area waveform (GAW, if accessible), we segmented the supra-glottal area waveform (SAW) based on high-speed imaging data. We defined SAW as the lateral contour of the vibrating supraglottal structure, e.g., ventricular folds, arytenoids, and epiglottis.
To evaluate time-dependent non-linear effects, we computed the phasegrams of GAW, SAW, audio, and EGG following the instructions in (14). Further, we analyzed the phase synchrony (PS) between EGG, SAW, and GAW, mainly following the guidelines given in (15).

Results:
For these ten voice qualities, we found physiological reasons to categorize them into three characteristic groups.

Group 1: "Vocal Fry" and "Grunt" show no/only subordinated supraglottic vibrations and were primarily based on irregular vocal fold vibrations. Conscious pitch control was not valid for these voice qualities. Vocal Fry was characterized by slightly irregular, long VF closing phases with a short intermediate opening. The supraglottis was adducted but no vibrations occurred here. In contrast, Grunt was a breathy/noisy quality with irregular VF vibrations with reduced VF adduction and disturbed oscillation sequence.

Group 2: “Distortion”, “Undertone”, “Rattle, and “Growl” are characterized by quasi-periodic movements of both the glottis and the supraglottis. The singer could consciously control his pitch and change it within a certain range. The ratio of glottal to supraglottal fundamental frequency (GSR) was 1:1 in Distortion and 2:1 in Undertone over the entire range. In Rattle, GSR changed from 4:1 to 7:1 during the glide, and in Growl from 2:1 to 3:1. While only parts of the ventricular fold vibrated in Distortion, full ventricular fold vibrations with complete closure were observed in Undertone and Rattle. In Growl, the entire epiglottis and the aryepiglottic fold also vibrated. In phasegrams, Undertone shows a bifurcation at the beginning of the quality. That is maintained within glissando. Distortion shows nearly no bifurcation throughout the whole task. Growl shows bifurcations with the onset and during the glide, while Rattle has a complex pattern.

Group 3: “Deathgrowl”, “Deathshout”, “Full Distortion”, and “Fry Scream” are characterized by extensive, irregular supraglottic vibrations of the entire epiglottis and the arytenoids. No subjective pitch control was possible for these qualities. Only in Deathshout, an additional quite regular glottal activity seemingly occurs.

Conclusions:

In the artistic field, differentiated, conscious control via supraglottic vibrations can be achieved. Pitch control was valid for four quasi-periodic vibration types, which can be differentiated in their relations between supra- vs. glottal vibrations. Entrainment phenomena of both oscillators were observed in sustained phonation. During the glissando, changes in synchronicity occurred frequently, which could be explained by a slight lag of one of the oscillators.

Acknowledgements:

The work of Louisa Traser and Johannes Fischer is funded by the DFG (TR1491/4-1 & FI 2803/1-1). The work of Matthias Echternach was funded by the DFG (Ec409/1-4).

References:

(1) Titze IR, Story BH. Acoustic interactions of the voice source with the lower vocal tract. 2014;101(4):2234–43.
(2) Maryn Y, De Bodt MS, Van Cauwenberge P. Ventricular dysphonia: Clinical aspects and therapeutic options. Laryngoscope. 2003;113(5):859–66.
(3) Kruse E, Michaelis D, Zwirner P, Bender E. Stimmfunktionelle Qualitatssicherung in der kurativen Mikrochirurgie der Larynxmalignome. HNO. 1997;45(9):712–8.
(4) Aaen M, McGlashan J, Sadolin C. Laryngostroboscopic Exploration of Rough Vocal Effects in Singing and their Statistical Recognizability: An Anatomical and Physiological Description and Visual Recognizability Study of Distortion, Growl, Rattle, and Grunt using laryngostroboscopic Imaging. J Voice. 2020;34(1):162.e5-162.e14.
(5) Güths RC, Rolim MRP, Coelho A. Glottal Voice Distortions: Nasolaryngoscopic and Spectral Analysis of Anatomophysiologic Changes in Singing Voice. Journal of Voice. 2021:38(1):31-39
(6) Guzman M, Barros M, Espinoza F, Herrera A, Parra D, Muñoz D, et al. Laryngoscopic, acoustic, perceptual, and functional assessment of voice in rock singers. Folia Phoniatr Logop. 2013;65(5):248–56.
(7) Bailly L, Henrich N, Pelorson X. Vocal fold and ventricular fold vibration in period-doubling phonation: Physiological description and aerodynamic modeling. J Acoust Soc Am. 2010 May;127(5):3212–22.
(8) Bergevin C, Narayan C, Williams J, Mhatre N, Steeves J, Bernstein JGW, et al. Overtone focusing in biphonic Tuvan throat singing. Elife. 2020 Feb 1;9.
(9) Zangger Borch D, Sundberg J, Lindestad PÅ, Thalén M. Vocal fold vibration and voice source aperiodicity in “dist” tones: A study of a timbral ornament in rock singing. Logop Phoniatr Vocology. 2004;29(4):147–53.
(10) Fuks L, Hammarberg B, Sundberg J. A self-sustained vocal-ventricular phonation mode: acoustical, aerodynamic and glottographic evidences. Tmh-Qpsr. 1998;3:49–59.
(11) Whitehead RL, Metz DE, Whitehead BH. Vibratory patterns of the vocal folds during pulse register phonation. 2014;75:1293–7.
(12) Lindestad PÅ, Blixt V, Pahlberg-Olsson J, Hammarberg B. Ventricular fold vibration in voice production: A high speed imaging study with kymographic, acousitc and perceptual analyses of a voice patient and a vocally health subject. Logop Phoniatr Vocology. 2004;29(4):162–70.
(13) Echternach M, Burk F, Köberlein M, Herbst CT, Döllinger M, Burdumy M, et al. Oscillatory Characteristics of the Vocal Folds Across the Tenor Passaggio. J Voice. 2017;31(3):381.e5-381.e14.
(14) Herbst CT, Herzel H, Švec JG, Wyman MT, Fitch WT. Visualization of system dynamics using phasegrams. J R Soc Interface. 2013;10(85).
(15) Pedersen M, Omidvarnia A, Zalesky A, Jackson GD. On the relationship between instantaneous phase synchrony and correlation-based sliding windows for time-resolved fMRI connectivity analysis. Neuroimage. 2018;181:85–94.

DETECTION OF COMMON VOICE PATHOLOGIES BY PRINCIPAL COMPONENT ANALYSIS OF VOICE ACOUSTIC PARAMETERS

Lady Catherine Cantor-Cutiva[1], Aishwarya Ramani[2], Eric J. Hunter[1]

[1] Department of Communication Sciences and Disorders, The University of Iowa, IA, USA
[2] Department of Communicative Sciences and Disorders, Michigan State University, MI, USA

Keywords: **Acoustic Parameters, Principal Component Analysis, Voice Assessment, Screening**

Abstract:

Background: The Perceptual Voice Qualities Database (PVQD) comprises 296 high-quality audio recordings with diverse speech samples, including sustained vowels /a/ and /i/, and sentences from the Consensus Auditory-Perceptual Evaluation of Voice. This database contains a wide spectrum of voice quality, covering various severities, speaker ages, and sexes. **Objective**: Identify diverse types of voice pathologies by analyzing key components of voice acoustic metrics, including jitter, shimmer, Harmonics-to-Noise Ratio (HNR), Pitch Period Entropy (PPE), CPP, and Alpha ratio, to expedite voice assessment. **Methods**: We used Principal Component Analysis (PCA) for exploratory data analysis and predictive modeling. PCA reduces data dimensionality by projecting data points onto the most influential principal components, thereby preserving critical data variation. We conducted voice acoustic analysis on the 296 PVQD voice samples using custom MATLAB scripts and PRAAT. The metrics included jitter, shimmer, HNR, CPP, CPPS, Alpha ratio, and PPE, among other metrics. **Results**: Our PCA findings revealed that CPP, shimmer, HNR, and PPE jitter were the components with the highest explanatory power for distinguishing pathological voices from normal controls. Using Generalized Estimating Equations, we found a statistically significant increase in HNR and PPE in healthy voices, while shimmer was reduced. **Conclusion**: Our analysis highlights the significance of HNR, PEE, and shimmer as the most predictive parameters for discriminating between pathological and normal voices. This insight can significantly enhance the efficiency of voice assessment.

Acknowledgements:

The Voice Foundation's Advancing Scientific Voice Research Grant Mechanism funded the collection of the PVQD recordings and online shareable access (P.I. Patrick R. Walden). The data analysis was supported by the National Institute of Deafness and Other Communication Disorders of the National Institutes of Health under Award Number R01DC012315 (P.I. Eric Hunter). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References:

de Abreu SR, da Silva Sousa ES, de Moraes RM, Lopes LW. Performance of acoustic measures for the discrimination among healthy, rough, breathy, and strained voices using the feedforward neural network. J Voice. 2022

Englert M, v Latoszek BB, Behlau M. Exploring the validity of acoustic measurements and other voice assessments. J Voice. 2022

Walden PR. Perceptual voice qualities database (pvqd): Database characteristics. J Voice. 2022;36(6):875. e15-875. e23

OBSERVATION OF LARYNGEAL BEHAVIORS VIA HIGH-SPEED VIDEOENDOSCOPY ASSOCIATED WITH THE PRIMARY SIGNS OF ADLD

Katherine L. Marks[1], Jennifer M. Vojtech[1], Manuel E. Díaz-Cádiz[1], Taylor F. Feaster[1], Bonnie Little[1], Giavanna Siracusano[1], Alex Estrada[1], Jose Rojas Olavarria[1], Daniel P. Buckley[1,2], Pavan Mallur[3], Gregory Grillone[2], J. Pieter Noordzij[2], Lauren Tracy[2], Cara E. Stepp[1,2,4]

[1]Speech, Language, and Hearing Sciences, Boston University, Boston, MA, USA
[2]Otolaryngology – Head and Neck Surgery, Boston University Medical School, Boston, MA, USA
[3]Otolaryngology– Head and Neck Surgery, Harvard University Medical School, Boston, MA
[4]Biomedical Engineering, Boston University, Boston, MA, USA

***Keywords*: Voice; High-Speed Videoendoscopy; Acoustic Analysis; Laryngeal Dystonia**

Abstract:

Objectives / Introduction: Adductor laryngeal dystonia (AdLD) is a neurological voice disorder in which spasms of the laryngeal muscles occur during voicing. Laryngeal spasms are associated with irregularities in voice fundamental frequency called *discontinuities*, and occur acoustically as phonatory breaks, frequency shifts, or creak [1, 2]. Although videostroboscopic imaging of the laryngeal mechanism is one of the primary methods used in voice clinics to detect or rule out pathology, the technique cannot capture laryngeal spasms. This is because laryngeal spasms occur at a faster timescale than the frame rate cap of videostroboscopy [30 frames per second; 3]. As such, evidence of laryngeal spasms is either associated with relatively slow aberrant behaviors such as supraglottic compression [e.g., hyperadduction, false vocal fold constriction, laryngeal constriction; 4] or is interpreted from auditory-perceptual assessments or acoustic signal analyses.

High-speed videoendoscopy (HSV) has been applied in laboratory settings to study the supraglottic and vibratory aspects of voice production—such as the factors contributing to glottic insufficiency [1]—as well as components related to vocal tremor [5]. Specific to AdLD, evidence of supraglottic compression [5] and irregularities in vibratory components such as oscillatory breaks [2] have been observed from HSV during sustained phonations. Segments of vocal fold obstruction have also been observed as more prevalent during connected speech in speakers with than without AdLD [6, 7]; such obstructions can be due to any combination of the epiglottis, arytenoid cartilage(s), false vocal folds, and/or laryngeal constriction masking the true vocal folds. Although these initial findings show promise toward elucidating the pathophysiology of laryngeal spasms in AdLD, no study to date has examined laryngeal behaviors in speakers with AdLD relative to the acoustic discontinuities that co-occur during laryngeal spasms.

Thus, the purpose of this study was to describe the laryngeal behaviors that co-occur with acoustic discontinuities in in speakers with and without AdLD during speech. We hypothesized that the presence of acoustic discontinuities would be statistically significantly related to the presence of supraglottic compression and vocal fold obstructions. Moreover, we predicted a greater occurrence of supraglottic compression and vocal fold obstructions in speakers with AdLD compared to controls.

Methods: Simultaneous microphone and flexible nasendoscope recordings at 1000 fps were collected from speakers with (n=10) and without AdLD (controls; n=10) as they read aloud 14 complex sentences that were designed to elicit laryngeal spasms via maximizing the occurrence of voiced phonemes. Trained technicians manually labeled acoustic discontinuities in the microphone signal using Praat software [8]; the same technicians independently labeled these discontinuities, then met to form consensus labels for any label in which there was disagreement. Custom MATLAB software [3] was used to identify sources of aberrant laryngeal behaviors, including obstructions (epiglottis, arytenoid cartilages, false vocal folds, laryngeal constriction) or supraglottic compression from the HSV images. The same technicians independently labeled these behaviors, then met to form consensus labels for any label in which they disagreed.

A multinomial logistic regression was performed to examine the relationship between the presence of acoustic discontinuities—labeled as break, creak, frequency shift, or no discontinuity—and that of the aberrant laryngeal behaviors of vocal fold obstruction and supraglottic compression. Additional fixed factors included in the regression model were group (AdLD, control), sex, and the interactions of group × obstruction and group × supraglottic compression. Age was included as a covariate. Loglikelihood ratio tests were used to examine the effects of the predictors on acoustic discontinuities at a significance level of $p < .05$. *Post hoc* Chi-square tests of independence were performed to examine significant interaction effects at a significance level of $p < .05$.

Results: The overall model was found to be statistically significant ($X^2(21) = 12002$, $p < .001$), with McFadden's $R^2 = .59$. All factors were significant in predicting the odds of an acoustic discontinuity occurring, including: age ($X^2(3) =$

88.9, $p < .001$), sex ($X^2(3) = 263.8$, $p < .001$), group ($X^2(3) = 32.2$, $p < .001$), obstruction events ($X^2(3) = 1301.7$, $p < .001$), supraglottic compression events ($X^2(3) = 3931.2$, $p < .001$), and the interactions of group × obstruction events ($X^2(3) = 19.3$, $p < .001$) and group × supraglottic compression events ($X^2(3) = 10.8$, $p = .013$). These findings support our hypothesis that acoustic discontinuities are related to the presence of supraglottic compression and vocal fold obstructions. Interestingly, both types of aberrant behavior only occurred in the presence of an acoustic discontinuity.

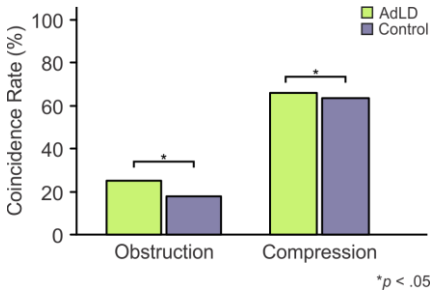


Figure 1. *Coincidence rate between aberrant behaviors and acoustic discontinuities.*

Vocal fold obstruction coincided with an acoustic discontinuity significantly more often in speakers with AdLD (25.4%) than controls (18.1%, $p < .001$; **Fig. 1**). Observations of supraglottic compression coincided with an acoustic discontinuity slightly more often for speakers with AdLD (66.3%) than controls (63.9%, $p < .001$). It is important to note that these aberrant laryngeal behaviors are not mutually exclusive. These results support our hypothesis that aberrant laryngeal behaviors coincide with acoustic discontinuities more often in individuals with AdLD than in controls.

Obstructions coincided with frequency shifts significantly more often in individuals with AdLD (20.9%) than in controls (6.4%, $p < .001$; **Fig. 2a**). The rate of coincidence between supraglottic compression events and acoustic discontinuities were similar across groups for both creak and frequency shifts; however, controls exhibited a substantially greater rate for phonatory breaks (90.0%) than individuals with AdLD (47.6%, $p = .011$; **Fig. 2b**). This surprising result may be due to the sensory effects of the nasoendoscope; further investigation is warranted.

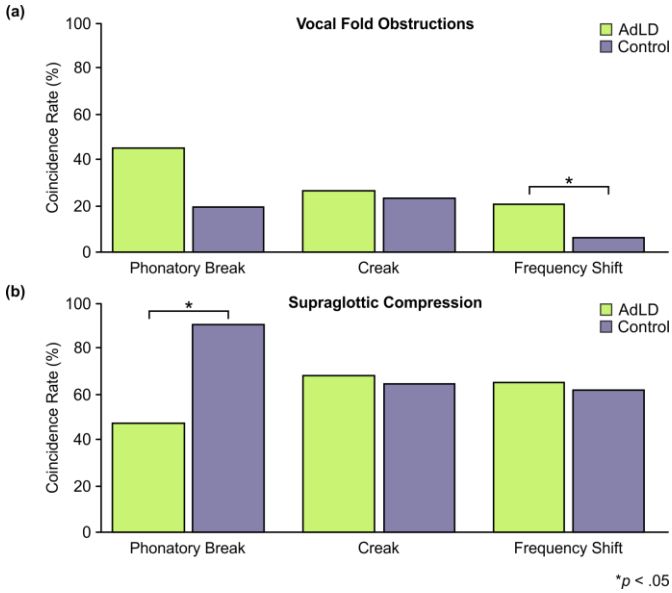


Figure 2. *Coincidence rate between aberrant behaviors and acoustic discontinuities, according to discontinuity type. Behaviors are shown as (a) vocal fold obstructions, and (b) supraglottic compression.*

Conclusions: The results of this study suggest that the manifestation of acoustic discontinuities differs between AdLD and typical voice, as we observed that aberrant laryngeal behaviors were more often present when a discontinuity occurred in speakers with AdLD. More importantly, our findings indicate that acoustic discontinuities involving phonatory breaks, sudden shifts in fundamental frequency, or creak may occur as the result of a vocal fold obstruction and/or instances of supraglottic compression. Further exploration of direction and extent of supraglottic compression is warranted.

Acknowledgements:

This work was supported by grant DC015570 (C.E.S), DC020349 (K.L.M), and DC020598 (K.L.M) from the National Institute on Deafness and Other Communication Disorders, a Speech Science Grant (K.L.M) from the American Speech-Language-Hearing Foundation, and a Student Research Grant (K.L.M.) from Sargent College, Boston University.

References:

1. Brown, W.S., Jr., H.B. Rothman, and C.M. Sapienza, *Perceptual and acoustic study of professionally trained versus untrained voices.* Journal of Voice, 2000. **14**(3): p. 301-309.
2. Marks, K.L., et al., *Spectral aggregate of the high-passed fundamental frequency and its relationship to the primary acoustic features of adductor laryngeal dystonia.* Journal of Speech, Language, and Hearing Research, 2022. **65**(11): p. 4085-4095.
3. Hillel, A.D., *The study of laryngeal muscle activity in Normal Human Subjects and in Patients With Laryngeal Dystonia Using Multiple Fine-Wire Electromyography.* 2001.
4. Leonard, R. and K. Kendall, *Differentiation of spasmodic and psychogenic dysphonias with phonoscopic evaluation.* The Laryngoscope, 1999. **109**(2): p. 295-300.
5. Parker, L.A., et al., *Reliability of high-speed videoendoscopic ratings of essential voice tremor and adductor spasmodic dysphonia.* Journal of Voice, 2019. **33**(1): p. 16-26.
6. Naghibolhosseini, M., et al., *Temporal segmentation for laryngeal high-speed videoendoscopy in connected speech.* Journal of Voice, 2018. **32**(2): p. 256. e1-256. e12.
7. Yousef, A.M., et al., *Detection of vocal fold image obstructions in high-speed videoendoscopy during connected speech in adductor spasmodic dysphonia: A convolutional neural networks approach.* Journal of Voice, 2022.
8. Boersma, P. and D. Weenink, *Praat: Doing phonetics by computer*. 2013: Amsterdam, The Netherlands.

OBJECTIVE QUANTIFICATION OF SPASM SEVERITY & TREATMENT RESPONSE IN ADDUCTOR LARYNGEAL DYSTONIA WITH HIGH-RESOLUTION MANOMETRY

Jesse Hoffmeister, PhD[1], Jürgen Konczak, PhD[2], Stephanie Misono, MD, MPH[1]

[1] Department of Otolaryngology, University of Minnesota, Minneapolis, Minnesota, USA

[2] Department of Kinesiology, University of Minnesota, Minneapolis, Minnesota, USA

***Keywords*: Laryngeal Dystonia; Spasm Strength; High-Resolution Manometry; Laryngeal Biomechanics**

Abstract:

Objectives / Introduction: Adductor Laryngeal Dystonia (AdLD) is a task-specific disorder in which spasms of laryngeal muscles during speech cause voice disruption, leading to communication breakdown and degraded quality of life[1]. AdLD is typically treated with injection of botulinum neurotoxin (BTX) into the larynx. The benefit of BTX is transient, so injection must be repeated every 3-4 months for life[2]. BTX can also cause temporary but problematic side effects of weak voice and dysphagia. A challenge to measuring the therapeutic impact of AdLD treatments, including BTX, is that current assessment tools such as acoustic and aerodynamic assessment and listener perceptual ratings have limited sensitivity to treatment response, and are rarely correlated with one another[3,4]. This also hinders evaluation of novel treatment options. Therefore, there is a critical need for objective measurements that are sensitive to treatment response and reflect patient report of symptom severity. We present preliminary findings using a new method for quantifying strength of laryngeal adduction during speech as a means for objectively assessing treatment response: measurement of upper esophageal sphincter (UES) pressure with high resolution manometry (HRM) (Fig 1). The UES is immediately adjacent to the larynx, and pressure produced by the UES is temporally coupled to strength of vocal fold adduction[5,6]. The biological purposes of the UES include protection of the airway by contracting to prevent tracheal aspiration of extra-esophageal reflux during behaviors that increase intrathoracic pressure, such as vocal fold adduction. During vocal fold adduction, UES pressure increases and stronger adduction results in greater UES pressure[7–9]. Measuring UES pressure thus has potential to quantify spasm severity. Spasms in AdLD cause vocal fold hyperadduction leading to voice disruption, and are hypothesized to increase UES pressure relative to acoustically normal phonation in the absence of spasms. We tested the hypotheses that UES pressure would be greater with greater symptom severity, and that UES pressure during voice disruption would significantly decrease 1 month after BTX injection compared to immediately prior to injection

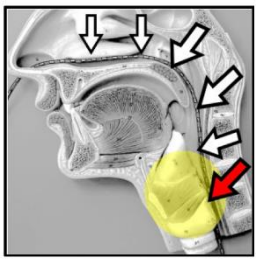


Fig 1 Sagittal view of HRM. White arrow point to HRM: catheter; yellow shading: larynx, adjacent to UES; red arrow: UES.

Methods:

Participants: Patients with AdLD who were known responders to BTX performed standardized speaking tasks during simultaneous HRM and voice recording at 2 time points: 1) when maximally symptomatic (within 1 week prior to BTX injection), and 2) when experiencing maximum benefit of BTX injection (1 month after injection)[10]. Patient-reported measures of symptom severity were collected at both time points using the Voice Handicap Index-10 (VHI-10), a 10-item questionnaire with a minimum score of 0 (minimal voice problem) and a maximum score of 40 (worst voice problem).

Procedure: After completion of patient-reported outcomes, a catheter with 36 circumferentially oriented sensors spaced at 1 cm intervals was passed transnasally to the proximal esophagus. Participants were given 5 minutes to acclimate to the catheter. Standardized speaking tasks were then recorded with an omnidirectional microphone positioned 2.5cm from the mouth, using a mobile pre-amplifier, and sampling rate of 41.5 kHz.

Data Analysis: Participants were categorized as strong or weak BTX responders using an established minimal clinically important difference in VHI10 score change of 6 as a cutoff[11]. Episodes of a common voice disruption associated with laryngeal spasms (vocal creak)[12] were identified using an automated MATLAB algorithm[13] and time-aligned with HRM pressure data for analysis (Fig 2). The average peak UES pressure increase from resting pressure during all instances of voice disruption was obtained. A linear mixed effects model assessed the influence of interactions between treatment time point (pre-, and 1 month post-BTX injection) and BTX treatment responsiveness (strong or weak) on percent increase in peak UES pressure during voice disruptions. Relationships between UES pressures during voice disruptions and VHI-10 scores were further assessed using repeated measures correlation[14] with all participants regardless of degree of change of VHI-10 after BTX.

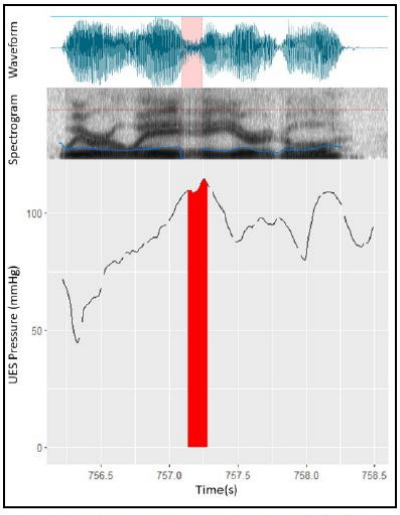


Fig 2 UES pressure and vocal acoustics in patient with AdLD saying, "We were away a year ago". Top panels show waveform and spectrogram, bottom panel shows UES pressure. Pink and red shading show voice disruption. Note sharp increase in UES pressure during voice

Results: Nine participants (mean age: 52.4 years, range 31-71 years) were recruited. Four were strong BTX responders (mean VHI-10 change=-13.3, SD 2.3) and 5 were weak responders (mean VHI-10 change=-0.6). There was a significant positive repeated measures correlation between VHI-10 scores and UES pressure during voice disruption (r_{rm}=0.68, 95%CI=0.08-0.91, p=0.03) (Fig3). There was also a significant interaction between treatment time point (pre- and 1 month post-BTX injection) and BTX treatment responsiveness on UES pressure during voice disruption (F(1,7)=6.21, p=0.04. In post-hoc testing, strong responders had a significant reduction in UES pressure during voice disruption after BTX (estimated mean difference=107.3%, SE=32.8% p=0.01), with large effect size (Cohen's *d*=1.2, 95% CI=0.12-2.29), while weak responders had almost no change in UES pressure during voice disruption after BTX (estimated mean difference=2.22%, SE=29.3%, p=0.94) (Fig4).

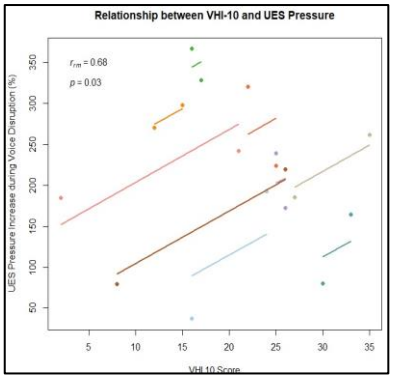


Fig 3: Relationship between UES pressure and VHI-10. Color represents individual participants and colored lines represent repeated measures correlation fit for each participant.

Conclusions: Measuring UES pressure during phonation could be used to objectively quantify spasm severity and thus treatment response in adductor laryngeal dystonia. We observed that UES pressure peaks during voice disruption were positively correlated with VHI-10 scores. Further, UES pressure peaks during voice disruptions decreased significantly after treatment with botulinum toxin, but only in individuals with a meaningful improvement in voice symptoms as measured with the VHI-10. This study is limited by the small sample size. In addition, we only assessed UES pressures during a common type of voice disruption in laryngeal dystonia (creak), and not in other voice disruptions such as phonation breaks or pitch shifts. Future work will determine if these findings can be replicated in larger samples. In addition, we will explore relationships between voice disruption and other measures that are possible with HRM during phonation, including duration of UES pressure spikes during voice disruption, as well as UES pressure measures during other types of voice disruption in AdLD (phonation breaks and pitch shifts as well as creak). The positive correlation between VHI-10 scores and UES pressure peaks during voice disruption suggest that this methodology could be used as a tool to quantify and to characterize spasm severity in this population, particularly for assessment of disease severity and treatment response.

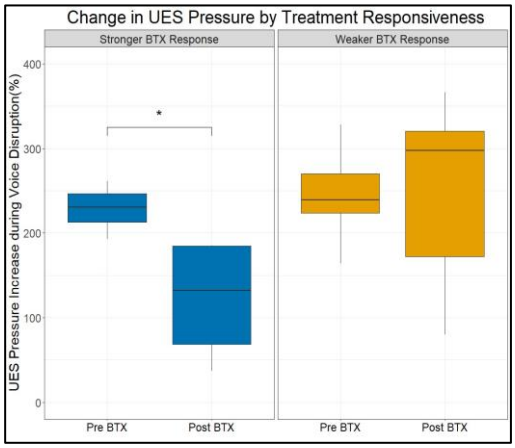


Fig 4: The change in UES pressure with BTX treatment depends upon whether an individual had a stronger or weaker response as measured by the minimal clinical important difference on the VHI-10.

Acknowledgements: We would like to acknowledge Ali Stockness, MS, Ashley Ramaker, MS, Maisie Simpson, BS, Augie Richter, and Joseph Hayek research coordination.

References:

1. Baylor CR, Yorkston KM, Eadie TL. The consequences of spasmodic dysphonia on communication-related quality of life: A qualitative study of the insider's experiences. *J Commun Disord*. 2005;38(5):395-419. doi:10.1016/j.jcomdis.2005.03.003
2. Faham M, Ahmadi A, Silverman E, Harouni GG, Dabirmoghaddam P. Quality of Life After Botulinum Toxin Injection in Patients With Adductor Spasmodic Dysphonia; a Systematic Review and Meta-analysis. *J Voice*. 2021;35(2):271-283. doi:10.1016/j.jvoice.2019.07.025
3. Simonyan K, Barkmeier-Kraemer J, Blitzer A, et al. Laryngeal Dystonia: Multidisciplinary Update on Terminology, Pathophysiology, and Research Priorities. *Neurology*. 2021;96(21):989-1001. doi:10.1212/WNL.0000000000011922
4. Rumbach A, Aiken P, Novakovic D. Treatment Outcome Measures for Spasmodic Dysphonia: A Systematic Review. *J Voice*. Published online May 3, 2022. doi:10.1016/j.jvoice.2021.10.001
5. Perera L, Kern M, Hofmann C, et al. Manometric evidence for a phonation-induced UES contractile reflex. *Am J Physiol - Gastrointest Liver Physiol*. 2008;294:885-891. doi:10.1152/ajpgi.00470.2007.
6. Meisoll FJ, Jungheim M, Fast JF, Miller S, Ptok M. Upper Esophageal Sphincter Response to Laryngeal Adductor Reflex Elicitation in Humans. *The Laryngoscope*. 2021;131:1778-1784. doi:10.1002/lary.29166
7. Amaris M, Dua KS, Naini SR, Samuel E, Shaker R. Characterization of the upper esophageal sphincter response during cough. *Chest*. 2012;142(5):1229-1236. doi:10.1378/chest.12-0638
8. Peters K, Miller S, Ptok M, Jungheim M. Phonation-induced Upper Esophageal Sphincter Contraction Caused by Different Phonation Types. *J Voice*. Published online 2022. doi:10.1016/j.jvoice.2022.06.006
9. Vaiano T, Herbella FAM, Behlau M. Pharyngeal , upper esophageal sphincteric and esophageal pressures responses related to vocal tasks at the light of high resolution manometry. *Arq Gastroenterol*. 2021;58(3):296-301.
10. Sapienza CM, Cannito MP, Murry T, Branski R, Woodson G. Acoustic Variations in Reading Produced by Speakers With Spasmodic Dysphonia Pre-Botox Injection and Within Early Stages of Post-Botox Injection. *J Speech Lang Hear Res*. 2002;45(5):830-843. doi:10.1044/1092-4388(2002/067)
11. Misono S, Yueh B, Stockness AN, House ME, Marmor S. Minimal Important Difference in Voice Handicap Index–10. *JAMA Otolaryngol Neck Surg*. 2017;143(11):1098-1103. doi:10.1001/jamaoto.2017.1621
12. Marks KL, Feaster TF, Baker S, D íaz CME, Doyle PC, Stepp CE. Spectral Aggregate of the High-Passed Fundamental Frequency and Its Relationship to the Primary Acoustic Features of Adductor Laryngeal Dystonia. *J Speech Lang Hear Res*. 2022;65(11):4085-4095. doi:10.1044/2022_JSLHR-22-00157
13. Kane J, Drugman T, Gobl C. Improved automatic detection of creak. *Comput Speech Lang*. 2013;27(4):1028-1047. doi:10.1016/j.csl.2012.11.002
14. Bakdash JZ, Marusich LR. Repeated measures correlation. *Front Psychol*. 2017;8(MAR):1-13. doi:10.3389/fpsyg.2017.00456

RELATION BETWEEN VOICE ONSET TIME, VOCAL HYPERFUNCTION TYPE AND VOICE QUALITY: AN EXPLORATORY STUDY

Maria Francisca de Paula Soares[1,2], Marilia Sampaio[1,2], Meike Brockmann-Bauser[2]

[1] Department of Speech, Language and Hearing Science, Federal University of Bahia, Salvador 40110-170 Brazil
[2] Division of Phoniatrics and Speech Pathology, University Hospital Zurich, University of Zurich, Zurich, Switzerland

***Keywords*: Acoustic Voice Analysis, Voice Onset Time (VOT), Vocal Hyperfunction, Voice Quality**

Abstract:

Objectives / Introduction:

Deviations in human voice onset are frequently described perceptual characteristics of vocal dysfunction. Objective measures for the onset of vocalization, such as Voice Onset Time (VOT), have been shown to be influenced by vocal pathology. However, it is unclear whether changes in VOT are more related to changes in perceptual voice quality, the degree of vocal hyperfunction or structural changes. Moreover, it has been shown that prosody and coarticulation related variables such as speaking fundamental frequency and intensity, following vowel and syllable stress affect VOT in vocally healthy individuals [1-4]. So far, the interrelation between these physiologic and voice pathology related influencing factors has not been investigated in a voice disordered population.

Therefore, main aim of the present work was to investigate in adults with non-phonotraumatic vocal hyperfunction or phonotraumatic vocal hyperfunction if a) perceptual voice quality and the type of vocal dysfunction are related to VOT, and if b) VOT is affected by the variables speaking fundamental frequency and speaking voice intensity, syllable stress and following vowel.

Methods:

In an exploratory retrospective study, 30 adults, of these 19 women with a mean age of 46.1 (SD 13.7) and 11 men with a mean age of 47.5 (SD 11.0) years, diagnosed with either non-phonotraumatic vocal hyperfunction (NPVH) or phonotraumatic vocal hyperfunction (PVH) were investigated. For perceptual analysis of voice quality, recordings of Cape-V sentences were assessed by one rater blinded to the diagnoses. Rated were the overall severity of dysphonia (OS), roughness, breathiness and strain on a 100 mm VAS scale. Assessment results were later conversed into the categories no (0-9), mild (10-34), moderate (35-69) and severe (> 70) voice deviation. To determine VOT, four samples of syllables with [p] plus vowel or diphthong were retrieved from the CAPE-V recordings. Acoustic analysis was conducted with Praat and comprised VOT, mean fundamental frequency (mean *fo*) and intensity (SPL dB(A)), and the coefficient of variation of fundamental frequency (CV_ *fo* %). Moreover, the phonetic variables following vowel and syllable stress were noted per each syllable.

Multivariate analysis of variance (MANOVA) was used to assess the effects of VH condition (NPVH vs. PVH) and OS (no/mild/moderate/severe) on VOT, *fo*, CV_ *fo* % and SPL, with Pillai's Trace for interpretation of results. Squared partial curvilinear correlation (η^2_p) was applied for assessing effect sizes. To investigate the relation between VOT, *fo* , SPL, and CV_ *fo* % with OS, Spearman's rank correlation coefficient was applied. The effects of the phonetic variables vowel context [[p] followed by e, I, O or aj]) and syllable stress (stressed vs. unstressed) on VOT, *fo* (mean and CV), and SPL were assessed by MANOVA. In all tests, gender was included as a covariable.

Results:

There was a highly significant influence of the overall severity of dysphonia on VOT, with a small effect size (p= <.001, η_p^2=.09). There were weakly positive and highly significant correlations between VOT and OS (r= 0.34, p≤.001), roughness (r=.29, p≤.001), breathiness (r=.22, p≤.001), and strain (r=.29, p≤.001). The post hoc Bonferroni test showed that VOT discriminated mild overall voice deviation (M=16.67) from moderate (M=21.62, p≤.001), and mild (M=16.67) from severe voice deviation (M=24.51, p≤.001). VH condition only affected CV_ *fo* (p=.02, η_p^2=.04), but not VOT, *fo*, and SPL (all p>.05). All voice quality subcharacteristics roughness, breathiness, and strain correlated significantly but weakly with mean speaking SPL (p≤.001), and not with mean *fo* (p>.05). Vowel context and syllable stress had no significant effect on VOT, *fo*, SPL, and CV_ *fo* (p>.05). Gender effects were only found for mean *fo* (p≤.001) and SPL (p=.01).

Conclusions:

In the present study of adults with non phonotraumatic and phonotraumatic hyperfunction, voice onset time was related to overall voice quality. In addition, with VOT it was possible to statistically discriminate mild from moderate and severe perceptual voice deviation. The results indicated a longer VOT in more severe dysphonia, but also in increased roughness, breathiness and strain. Thus, changes in voice onset time may be a main marker of audible dysphonia, regardless of the underlying perceptual characteristic. Notably, all perceptual characteristics were significantly related with speaking SPL, but not with fundamental frequency.

In our sample, VOT was not related to vocal hyperfunction type and thus the presence of secondary structural pathology associated with prolonged vocal misuse. Also, the speech related characteristics vowel context and syllable stress had no systematic influence on VOT. To further explore the diagnostic potential of VOT, the present results should be confirmed in a larger clinical study in women and men of different age groups, stratified after the degree of perceptual voice deviation and diagnosis type.

Acknowledgements:

We gratefully acknowledge all patients, students, and professors from the Department of Speech, Language and Hearing Science, Federal University of Bahia, Salvador who contributed to the database used in the present study.

References:

1. Colletti, L. and E. Heller Murray, *Voice Onset Time in Children With and Without Vocal Fold Nodules.* J Speech Lang Hear Res, 2023. 66(5): p. 1467-1478.
2. Groll, M.D., S. Hablani, and C.E. Stepp, *The relationship between voice onset time and increase in vocal effort and fundamental frequency.* Journal of Speech, Language, and Hearing Research, 2021. 64(4): p. 1197-1209.
3. McKenna, V.S., et al., *Voice Onset Time in Individuals With Hyperfunctional Voice Disorders: Evidence for Disordered Vocal Motor Control.* J Speech Lang Hear Res, 2020. 63(2): p. 405-420.

Cho, T. and P. Ladefoged, *Variation and universals in VOT: evidence from 18 languages.* Journal of Phonetics, 1999. 27(2): p. 207-229.

INTEGRATING BIOMECHANICAL MODELS, AMBULATORY MONITORING, AND STATISTICAL ANALYSIS FOR UNDERSTANDING NON-PHONOTRAUMATIC VOCAL HYPERFUNCTION

Carlos Calvache[1, 2, 3], Jesús Parra[4], Leonardo Solaque[3], Matías Zañartu[4]

[1] Department Communication Sciences and Disorders, Corporación Universitaria Iberoamericana, Bogotá, Colombia
[2] Vocology Research, Vocology Center, Bogotá, Colombia
[3] Department of Mechatronics Engineering, Universidad Militar, Bogotá, Colombia
[4] Department of Electronic Engineering, Universidad Técnica Federico Santa Maria, Valparaíso, Chile

***Keywords*: Non-phonotraumatic Vocal Hyperfunction, Vocal Therapy, Biomechanical Modeling, Ambulatory Monitoring**

Abstract:

Objectives / Introduction: Non-Phonotraumatic Vocal Hyperfunction (NPVH) is clinically manifested as muscle tension dysphonia, a specific subset of voice disorders marked by laryngeal muscular imbalance [1]. Biomechanical models like the asymmetric version of Triangular Body-Cover Model (A-TBCM) [2] has the potential to advance NPVH analysis, by offering precise simulations of its complex vocal dynamics, notably muscular imbalances and laryngeal asymmetries. The use of synthetic A-TBCM-generated data has opened new avenues for training neural networks to predict parameters, including muscle activations and subglottal pressure [3]. Incorporating tools like the Voice Health Monitor (VHM) and Impedance-Based Inverse Filtering (IBIF) is vital for estimating flow parameters in ambulatory settings [4]. Integrating these components enables exploration of novel approaches to NPVH analysis in ambulatory environments. This study aims to integrate biomechanical models like A-TBCM, neural networks, and ambulatory monitoring techniques for statistical NPVH analysis, for improving our understanding of this voice disorder.

Methods: Ambulatory voice monitoring involved 54 women, including 27 with NPVH and 27 in a matched control group. The monitoring spanned 7 days, with 15 NPVH subjects having an extra 7 days for pre- and post-therapy data. Acoustic and aerodynamic data were gathered and processed using VHM and IBIF, covering parameters like SPL, fo, H1-H2, ACFL, MFDR, CPP, SQ, and OQ [4]. Additionally, a neural network pretrained with the A-TBCM predicted Subglottal Pressure (PS) and intrinsic laryngeal muscle activations. Data preprocessing utilized interquartile ranges to detect and remove outliers, by considering typical ranges for each vocal feature in NPVH and control groups. Aerodynamic data (MFDR, ACFL, OQ, SP) were log normalized relative to SPL, following prior in laboratory methods [5]. This logarithmic normalization improved proportional scaling, facilitating meaningful vocal efficiency comparisons. Data analysis involved descriptive and inferential statistics, encompassing Statistical Moments (Mean, median, mode, standard deviation, percentiles, skewness, and kurtosis) and Multivariate Gaussians. Multivariate Gaussians offered a joint representation of trends and relationships among different vocal function measures, obtaining hyperparameters such as means, variances, and covariances. Subsequently, statistical tests were used to assess the validity of our findings, including the Shapiro-Wilk test for distribution normality, T-Test (T-T), Mann-Whitney test (M-W) for group differences, and Cohen's Distances for effect size measurement.

Results: Table 1 shows the most significant features resulting from comparative tests between NPVH subjects and the control group. NPVH patients exhibit skewed distributions in parameters like CPP and MFDR at the 25th percentile, suggesting heightened aerodynamic demand for equivalent vocal loudness. This implies increased vocal fold stiffness or tension and elevated aerodynamic effort. The negative CPP Skew, indicating higher NPVH values, accentuates vocal dynamics irregularity and potentially muscular imbalance. Furthermore, kurtosis in CPP distribution for NPVH patients reveals prominent peaks, highlighting glottic oscillation variability, potentially linked to altered vocal mechanics.

FEATURE	M-W P-VALUE	T-T P-VALUE	COHEN'S D
CPP P25	0.058400	--	0.502522
MFDR P25	--	0.048816	0.538727
CPP Skew	0.033830	--	-0.567629
CPP Kurtosis	--	0.039484	0.563993

Table 1. Differential features with significance between NPVH patients and control subjects

Table 2 presents results from analyzing the most significant features in comparative tests between NPVH subjects before and after vocal therapy. Notably, normalizing aerodynamic variables proved critical, by reducing the impact of external factors like environmental noise or individual vocal intensity, and resulting in improved vocal efficiency post-therapy. Normalized features are denoted with ' in Table 2. Increased glottal opening variability after therapy, measured via OQ' STD, suggests improved adaptability to diverse phonatory demands. Negative values indicate higher pre-vocal

therapy means. Changes in value distribution asymmetry, reflected in PS' and SPL skewness, signify enhanced aerodynamic control, adapting more effectively to vocal demands. Additionally, increased kurtosis in TA2 (TA with muscular imbalance) hints at reduced muscular imbalance among NPVH patient post-therapy. These findings emphasize the positive influence of vocal therapy on vocal production mechanics and control in NPVH patients.

FEATURE	M-W P-VALUE	T-T P-VALUE	COHEN'S D
OQ´ STD	0.044136	--	-0.323021
PS´ Skew	0.016254	--	-0.411390
PS Skew	0.025590	--	-0.727181
SPL Skew	--	0.054584	0.619700
TA2 Kurtosis	0.044136	--	-0.597228

Table 2. Differential features with significance between NPVH subjects pre and post voice therapy

Conclusions: This study combines biomechanical models, ambulatory monitoring techniques, and statistical analysis to understand NPVH. Results reveal significant differences in parameters like CPP and MFDR, indicating increased aerodynamic demand and vocal dynamics irregularity in NPVH. Analyzing NPVH patients pre- and post-vocal therapy shows improved vocal efficiency, adaptability to diverse phonatory demands, and reduced muscular imbalance. These findings hold important implications for NPVH diagnosis and treatment, affirming vocal therapy effectiveness in enhancing vocal function for affected individuals.

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 FB0008, and Corporación Universitaria Iberoamericana, Research Department, under search grant for project 202210D038. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References:

[1] M. Desjardins, C. Apfelbach, M. Rubino, and K. V. Abbott, “Integrative Review and Framework of Suggested Mechanisms in Primary Muscle Tension Dysphonia,” *Journal of Speech, Language, and Hearing Research*, vol. 65, no. 5, pp. 1867–1893, Apr. 2022, doi: 10.1044/2022_JSLHR-21-00575.

[2] J. A. Parra, C. Calvache, and M. Zañartu, “AN ASYMMETRIC TRIANGULAR BODY-COVER MODEL WITH UNBALANCED INTRINSIC MUSCLE ACTIVATION,” Phoenix, Arizona, Mar. 2023. Accessed: Jan. 30, 2024. [Online]. Available: https://ce.mayo.edu/sites/default/files/media/2023-03/AQL-2023_Final%20Podium%20Abstracts.pdf

[3] E. J. Ibarra *et al.*, “Estimation of Subglottal Pressure, Vocal Fold Collision Pressure, and Intrinsic Laryngeal Muscle Activation From Neck-Surface Vibration Using a Neural Network Framework and a Voice Production Model,” *Front Physiol*, vol. 12, no. September, pp. 1–13, 2021, doi: 10.3389/fphys.2021.732244.

[4] D. D. Mehta *et al.*, “Using ambulatory voice monitoring to investigate common voice disorders: Research update,” *Front Bioeng Biotechnol*, vol. 3, no. OCT, pp. 1–14, 2015, doi: 10.3389/fbioe.2015.00155.

[5] V. M. Espinoza, D. D. Mehta, J. H. Van Stan, R. E. Hillman, and M. Zañartu, “Glottal aerodynamics estimated from neck-surface vibration in women with phonotraumatic and nonphonotraumatic vocal hyperfunction,” *Journal of Speech, Language, and Hearing Research*, vol. 63, no. 9, pp. 2861–2869, 2020, doi: 10.1044/2020_JSLHR-20-00189.

ADDING VIBRATO-BASED AMBULATORY SINGING MEASURES IN THE CLINICAL ASSESSMENT OF PATIENTS WITH VOCAL HYPERFUNCTION

Emma C. Willis[1,2], Ben Kevelson[2], Jarrad H. Van Stan[1–3], Matías Zañartu[4], 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
[4] Department of Electronic Engineering, Universidad Técnica Federico Santa María, Valparaíso, Chile

Keywords: Singing voice; Ambulatory monitoring; Signal processing; Voice assessment

Abstract:

Objectives / Introduction: The prevalence of voice disorders in singers is estimated to be 46% [1]. While singers represent a rather small portion (~1%) of the overall workforce in the United States, they make up an estimated 11–29% of patients seeking treatment at voice centers [2, 3]. Despite this high prevalence, there is no standardized, quantitative protocol for the clinical assessment of the singing voice. Baker et al. [4] recently investigated aeroacoustic analysis of singing, including two newer vibrato-based perturbation measures in vocally healthy, professional classical singers in a controlled laboratory setting. $\text{Vibrato}_{\text{Jitter}}$ and $\text{Vibrato}_{\text{Shimmer}}$ measure the degree of perturbation in the frequency and amplitude of the vibrato (musical pitch modulation) during a sung sample [4]. Hunter and Schloneger's study [5] investigated relationships between vocal dose and voice quality, both observed and perceived, during the typical daily voice use of vocally healthy college-age singers. This study seeks to build off this previous work by applying similar aeroacoustic analysis to ambulatory voice monitoring data in professional singers with diagnosed vocal hyperfunction (either phonotraumatic vocal hyperfunction or non-phonotraumatic vocal hyperfunction) and vocally healthy matched controls for comparison. Singing-related measures could not only provide valuable assessment tools for singers, but also allow for objective assessment of treatment effects and biofeedback targets.

Methods: Data collected during an ongoing study were utilized, selected with the following criteria: participants were diagnosed with vocal hyperfunction, self-identified occupation being a professional singer, voice teacher, or voice student, and the data already passed quality checks. The data were processed with a singing voice detector algorithm that isolated singing voice segments for each day of ambulatory monitoring [7]. Using Audacity 3.3.0, the isolated singing voice WAV files were manually processed to only include singing with vibrato for vibrato analysis utilizing BioVoice [8]. BioVoice is a software tool for acoustic analysis of voice that estimates numerous acoustical parameters, including singing-specific vibrato rate, vibrato extent, $\text{Vibrato}_{\text{jitter}}$ and $\text{Vibrato}_{\text{shimmer}}$, and other measures such as fundamental frequency and formant frequencies. MATLAB code was utilized to add 100 ms of silence between singing segments to create clear voiced segments for the analysis of vibrato in BioVoice. This study investigates $\text{Vibrato}_{\text{Jitter}}$ and $\text{Vibrato}_{\text{Shimmer}}$ as well as several additional measures, including cepstral peak prominence, vocal efficiency [9], and estimates of glottal airflow characteristics that are correlated with increased vocal loading [10], which can be associated with vocal hyperfunction [6].

Participants are recruited through sequential convenience sampling and snowball sampling. 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 that included the collection of a complete case history, endoscopic imaging of the larynx, completion of the Voice-Related Quality of Life questionnaire, an auditory-perceptual evaluation using the Consensus Auditory-Perceptual Evaluation of Voice, and aerodynamic and acoustic assessments of vocal function. Each participant diagnosed with either phonotraumatic or nonphonotraumatic vocal hyperfunction was matched with a vocally heathy control based on age (±5 years), sex, occupation, and singing genre. The vocally healthy status of all matched-control candidates was verified via interview, auditory-perceptual evaluation by a speech-language pathologist, and a laryngeal stroboscopic examination. During the interview, each vocally healthy control participant was asked if they had any voice difficulties that had an impact on their daily life. If the matched-control candidate indicated voice difficulties or demonstrated a dysphonic vocal quality, they were excluded from further participation without completion of laryngeal stroboscopic examination.

Results and Discussion: Pilot analysis of a patient with nonphonotraumatic vocal hyperfunction showed daily variability across episodes of singing, yielding average pre-to-post treatment measures of $\text{Vibrato}_{\text{Jitter}}$ (13.8% to 12.7%), $\text{Vibrato}_{\text{Shimmer}}$ (37.8% to 60.9%), cepstral peak prominence (21.6 dB to 19.1 dB), and maximum flow declination rate (465.9 L/s^2 to 108.6 L/s^2). Although $\text{Vibrato}_{\text{Shimmer}}$ was expected to decrease post treatment, like the other measures, it was found to increase for this patient. Results will be presented on a larger cohort of singers with diagnosed vocal hyperfunction and their matched controls for comparison. Even though data from ambulatory monitoring are more ecologically valid because it samples daily vocal function, it does not require participants to produce a standardized set

of singing tasks. This could introduce a source of variability that would make comparisons between and within (pre- vs. post-treatment) participants more challenging. Also, in the future, our laboratory data collection protocol could include standardized singing tasks that could potentially be used to facilitate more controlled comparisons and enhance the interpretation of ambulatory measures.

Conclusions: Aeroacoustic measures, cepstral peak prominence, $Vibrato_{jitter}$, $Vibrato_{shimmer}$, and vocal efficiency ratios may be clinically useful in the ambulatory assessment of singing voice during daily life, particularly for patients presenting with more subtle voice complaints, as well as in objective assessment of treatment effects. Future research should further investigate these measures in professional singers with vocal hyperfunction and their vocally healthy matched controls using both ambulatory monitoring of daily vocal function and laboratory-based assessments that utilize well-controlled standardized singing tasks.

Acknowledgements:

This research was funded by the National Institutes of Health National Institute on Deafness and Other Communication Disorders (grants: R33 DC011588, P50 DC015446, and R01 DC019083). The contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH. The authors would like to thank Laura Toles and Katie Marks for study participant recruitment and data collection, MGH Voice Center research assistants for data pre-processing, AJ Ortiz for data analysis and signal processing, and Rob Petit and Dave Viggiano for mobile application development.

Drs. Robert Hillman and Daryush Mehta have a financial interest in InnoVoyce LLC, a company focused on developing and commercializing technologies for the prevention, diagnosis, and treatment of voice-related disorders. Dr. Matías Zañartu has a financial interest in Lanek SPA, a company focused on developing and commercializing biomedical devices. The financial interests of Drs. Hillman, Mehta, and Zañartu have been reviewed and are managed in compliance with the conflict-of-interest policies of Massachusetts General Hospital and Mass General Brigham, as well as Universidad Técnica Federico Santa Maria, respectively.

References:

[1] Pestana, P. M., Vaz-Freitas, S., Manso, M. C. (2017). Prevalence of Voice Disorders in Singers: Systematic Review and Meta-Analysis. Journal of Voice, 31(6):722-727.

[2] Hanschmann, H., Lohmann, A., and Berger, R. (2011). Comparison of subjective assessment of voice disorders and objective voice measurement. Folia Phoniatrica et Logopaedica, 63(2):83-87.

[3] Lin, J. Z., Espinoza, V. M., Marks, K. L., Zanartu, M., and Mehta, D. D. (2020). Improved subglottal pressure estimation from neck-surface vibration in patients with voice disorders. Proceedings of the International Conference on Voice Physiology and Biomechanics.

[4] Baker, C.P.; Purdy, S.C.; Rakena, T.O.; Bonnini, S. It Sounds like It Feels: Preliminary Exploration of an Aeroacoustic Diagnostic Protocol for Singers. J. Clin. Med. 2023, 12, 5130. https://doi.org/ 10.3390/jcm12155130

[5] Schloneger, M. J., & Hunter, E. J. (2017). Assessments of Voice Use and Voice Quality Among College/University Singing Students Ages 18-24 Through Ambulatory Monitoring With a Full Accelerometer Signal. Journal of voice: official journal of the Voice Foundation, 31(1), 124.e21–124.e1.2400000000001.24E32. https://doi.org/10.1016/j.jvoice.2015.12.018

[6] Rosenthal, A. L., Lowell, S. Y., & Colton, R. H. (2014). Aerodynamic and acoustic features of vocal effort. Journal of voice : official journal of the Voice Foundation, 28(2), 144–153. https://doi.org/10.1016/j.jvoice.2013.09.007

[7] Ortiz, A. J., Toles, L.E., Marks, K. L., Capobianco, S., Mehta, D. D., Hillman, R. E., Van Stan, J. H. (2019). Automatic speech and singing classification in ambulatory recordings for normal and disordered voices. J. Acoust. Soc. Am., 146 (1): EL22– EL27. https://doi.org/10.1121/1.5115804

[8] Morelli, M.S., Orlandi, S., Manfredi, C. (2021). BioVoice: A Multipurpose Tool for Voice Analysis. Biomed. Signal Process. Control 2021, 64, 102302 https://doi.org/10.1016/j.bspc.2020.102302

[9] Mehta, D. & Hillman, R. E. (2007). Use of aerodynamic measures in clinical voice assessment. SIG 3 Perspectives on Voice and Voice Disorders, 17, 14–18. https://doi.org/10.1044/vvd17.3.14

[10] M. Zañartu, J. C. Ho, D. D. Mehta, R. E. Hillman, and G. R. Wodicka, (2013) “Subglottal impedance-based inverse filtering of speech sounds using neck surface acceleration”, IEEE Trans. Audio Speech Lang. Proc., 21(9), pp. 1929-1939. DOI: 10.1109/TASL.2013.2263138

AUTOMATED CREAK DIFFERENTIATES LARYNGEAL DYSTONIA AND MUSCLE TENSION DYSPHONIA DURING A CONVERSATIONAL SPEECH TASK

Daria A. Dragicevic[1], Katherine L. Marks[1], Cara Sauder[2], Daniel P. Buckley[1,3], Laura E. Toles[4], Lauren F. Tracy[3], J. Pieter Noordzij[3], Gregory A. Grillone[3], Ted Mau[4], John Paul Giliberto[5], Tanya K. Meyer[5], Albert L. Merati[5], Tanya L. Eadie[2], Cara E. Stepp[1,3,6]

[1] Department of Speech, Language, and Hearing Sciences, Boston University, Boston, MA, USA Department of Speech and Hearing Sciences, University of Washington, Seattle, WA, USA
[2] Department of Otolaryngology-Head and Neck Surgery, Chobanian & Avedisian School of Medicine, Boston University, Boston, MA, USA
[3] Department of Otolaryngology-Head and Neck Surgery, Voice Center, University of Texas Southwestern Medical Center, Dallas, TX, USA
[4] Department of Otolaryngology-Head and Neck Surgery, University of Washington School of Medicine, Seattle, WA, USA
[5] Department of Biomedical Engineering, Boston University, Boston, MA, USA
***Keywords*: Voice; Laryngeal Dystonia; Vocal Hyperfunction; Creak**

Objectives/Introduction: The purpose of this study was to evaluate the discriminative ability of automated creak estimates in differentiating among speakers with adductor laryngeal dystonia (AdLD), speakers with muscle tension dysphonia (MTD), and speakers with no history of voice disorders during conversational speech. As the prevalence of creak previously showed promise in differentiating these groups using a standard reading passage, we built upon the previous work of Marks et al. (2023) to validate the performance in a larger sample size. Further, as literature suggests that typical speakers use more creak during conversational speech to increase inflection and expression (Agathe & Claire, 2013; Davidson, 2021; Wolk et al., 2012), we hypothesized that the automated creak estimates would be poorer at differentiating between groups when using conversational speech stimuli as compared to a standard reading passage.

Methods: Fifty speakers with AdLD, fifty speakers with MTD, and fifty speakers with no history of voice disorders, henceforth referred to as controls, were recorded reading the Rainbow Passage and producing a conversational speech sample. An opensource automated creak detector was used to calculate the percentage (%) of creak in the two stimuli types. A mixed models analysis of variance (ANOVA) was performed to determine the effects of group (between participants), stimuli type (within participants), and their interaction on the discriminative ability of % creak, with overall severity (OS) entered as a covariate. Receiver operator characteristic (ROC) curves were computed and the area under the curve (AUC) was evaluated for both stimuli types to determine the accuracy of the automated creak detector.

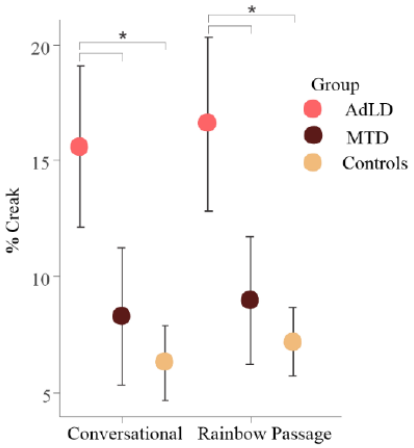


Figure 1: Mean % creak and 95% CI are plotted for each stimuli type (conversational speech and the Rainbow Passage) for each group (AdLD, MTD, and Controls). * = statistically significant difference at $p < .05$.

Results: The mean % creak and 95% confidence interval (CI) are displayed in Figure 1 for each stimuli type and group. Results of the ANOVA showed a statistically significant effect of group ($p < .001$) with a medium effect size ($\eta 2 = 0.05$), yet no significant effect of stimuli, interaction, nor OS. Post-hoc Wilcoxon rank sum tests were conducted and revealed statistically significant differences between the LD group and controls ($p < .001$), and statistically significant differences between the LD and MTD groups ($p < .001$). However, the post-hoc tests revealed no statistically significant difference between MTD and controls ($p = 0.71$).

Results of the ROC curve analyses for the conversational speech the Rainbow Passage are shown in Figure 2. The AUCs for discriminating between AdLD and MTD were .71 using conversational speech and .69 using the Rainbow Passage (see blue solid lines). The AUCs for discriminating between AdLD and Controls were .74 using conversational speech and .72 using the Rainbow Passage (see purple dotted lines). Finally, the AUCs for discriminating between Controls and MTD were .50 using conversational speech and .48 using the Rainbow Passage (see pink dashed lines).

Conclusions: The purpose of this study was twofold: to validate the discriminative ability of creak in AdLD, as found in Marks et al. (2023), in a larger sample size and to determine whether this discriminative ability would remain valid during a conversational speaking task. Although automated creak estimates provided statistically significant differences between individuals with AdLD and the two other groups at the group level, discrimination performance was lower than that found in previous work. This is reasonable given the global use of creak by all speakers and the large increase in sample size in the current study, supporting the generalizability of the current results.

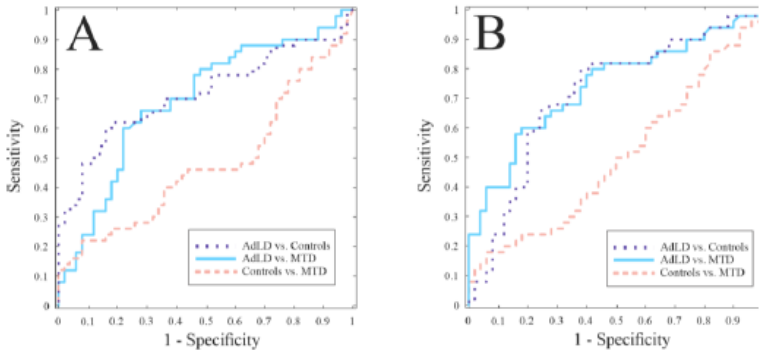


Figure 2: Receiver-operator characteristic curve plot illustrating the sensitivity and 1 - specificity of % creak in differentiating speakers with AdLD from Controls (purple dotted lines), AdLD from MTD (blue solid lines), and MTD from Controls (pink dashed lines). (A) ROC plot using the Rainbow Passage stimuli. (B) ROC plot using the conversational speech stimuli.

The discriminative ability indicated in the current study is on par with other previous acoustic metrics that require contrasts between specific stimuli (Houtz et al., 2010; Roy et al., 2014). In contrast, the ability of automated creak to differentiate speakers with AdLD from MTD and speakers with AdLD from controls was the same regardless of stimuli type: performance was similar in the conversational speaking task as the Rainbow Passage. Thus, automated creak estimates may have promise as one of several discriminative features and are not constrained by the requirement for collection of specific speech stimuli.

Acknowledgements:
This work was supported by the National Institutes of Health under grants DC015570 (CES) and DC013017 (CES), and by a Graduate Fellow Award from the Rafik B. Hariri Institute for Computing and Computational Science and Engineering (DAD). The authors thank Marlene Chavez-Corona and Courtney Dunsmuir for their assistance in preparing the data for this project.

References:
Agathe, B. L., & Claire, P. l. (2013). The Influence of language and speech task upon creaky voice use among six young American women learning French.
Davidson, L. (2021). The versatility of creaky phonation: Segmental, prosodic, and sociolinguistic uses in the world's languages. *WIREs Cognitive Science*, *12*(3), e1547. https://doi.org/https://doi.org/10.1002/wcs.1547
Houtz, D. R., Roy, N., Merrill, R. M., & Smith, M. E. (2010). Differential diagnosis of muscle tension dysphonia and adductor spasmodic dysphonia using spectral moments of the long-term average spectrum. *Laryngoscope*, *120*(4), 749-757. https://doi.org/10.1002/lary.20741
Marks, K. L., Díaz Cádiz, M. E., Toles, L. E., Buckley, D. P., Tracy, L. F., Noordzji, J. P., Grillone, G. A., & Stepp, C. E. (2023). Automated Creak Differentiates Adductor Laryngeal Dystonia and Muscle Tension Dysphonia. *The Laryngoscope*.
Roy, N., Mazin, A., & Awan, S. N. (2014). Automated acoustic analysis of task dependency in adductor spasmodic dysphonia versus muscle tension dysphonia. *The Laryngoscope*, *124*(3), 718-724. https://doi.org/https://doi.org/10.1002/lary.24362
Wolk, L., Abdelli-Beruh, N. B., & Slavin, D. (2012). Habitual Use of Vocal Fry in Young Adult Female Speakers. *Journal of Voice, 26*(3), e111-e116. https://doi.org/10.1016/j.jvoice.2011.04.007

DEVELOPMENT OF GERMAN CONTINUOUS SPEECH STIMULI FOR RELATIVE FUNDAMENTAL FREQUENCY (RFF)

Mark Berardi[1], Benjamin Rehring[1,2], Juliane von der Heyde[1], Hannah Moser[1], Maria Dietrich[1]

[1] Department of Psychiatry and Psychotherapy, University Hospital Bonn, Bonn, Germany
[2] Department of Neurology, University Hospital Bonn, Bonn, Germany

***Keywords*: e.g. Speech Acoustics; Multilingual; Voice Evaluation; Vocal Effort;**

Abstract:

Objectives / Introduction:

Vocal effort is a frequent complaint among those with voice disorders or occupational voice users. Relative fundamental frequency (RFF) is an acoustic measure used to estimate laryngeal tension and vocal effort. RFF measures changes in glottal cycles during the transition from voiced sonorants to voiceless consonants in speech. Both uniform utterances (e.g., /afa/) and continuous speech are used for RFF measurement. Previous continuous speech stimuli include prolonged speech passages (e.g., Rainbow Passage), CAPE-V sentences [1], and sentences developed specifically for measurement of RFF instances (English example: “The dew shimmered over my shiny blue shell again.” [2]). Currently, the only validated sentences containing targeted RFF segments are in English. While uniform utterances can be used in German, continuous speech samples better reflect natural voice production. As such, German RFF sentences are required to improve ecological validity. This study develops and evaluates novel German RFF sentences to establish stimuli for future vocal effort research. We hypothesize that at least one novel German RFF sentence for each fricative (/f/ and /ʃ/) will demonstrate token-level variability equal to or less than that of the validated English RFF sentences.

Methods:

For the development of German RFF sentences, a panel of three native German-speaking and academically-trained speech therapists developed potential RFF sentence stimuli in German. Based on the previous research, the target voiceless consonants with the best performance were /f/ or /ʃ/ [2]. Therefore, for each target voiceless consonant, three sentences were made to fit the same criteria as used in the development of English RFF sentence stimuli, which include:

(1) the sentence must be a statement and not a question,
(2) contain exactly 3 instances of the target fricative (/f/ or /ʃ/),
(3) contain no instances of the other target fricative,
(4) have the vowels surrounding the target fricatives be stressed,
(5) have the vowels surrounding the target fricative are not shwa vowels (/ə/)
(6) be relatively short (less than 20 syllables),
(7) be grammatically and semantically correct.

With the six sentences developed, 19 female participants (age: M: 27.5 years, SD: 3.0) were recorded to measure the variability of the RFF. The participants were all native German speakers with no current speech or hearing complaints. Each participant read the sentences and the uniform utterance triplet, /afa afa afa/, three times. The participants were recorded in a sound-treated clinic room with a head-mounted microphone (AKG C 520). Before RFF calculation, the target utterances were segmented using the web-based phonetic aligner, WebMAUS which uses the orthographic transcription of speech samples to segment the recording into the phonetic and word segments. Following the utterance segmentation, glottal pulses were manually extracted from the vowel offset and onset using Praat. From the extracted pulses, RFF was calculated as $RFF_n = 12 * \log_2(f_n/f_{ss})$, where n is the glottal cycle number from 1 to 10 with f_n as the instantaneous frequency of the n^{th} cycle and f_{ss} is the instantaneous frequency of the steady-state cycle. The mean token-level RFF standard deviation was computed to determine the least variable RFF sentence for each target voiceless consonant. English RFF speech samples from 19 age-matched females (age: M: 26.0 years, SD: 3.8) from a previous study were used as a comparison to the novel German stimuli [3].

Results:

Table 1 contains the six novel German sentences, the existing English sentences, and the uniform utterances with the computed mean token-level RFF standard deviations for the German-speaking participants and the age-matched English-speaking participants.

Table 1. Summary of mean token-level RFF standard deviations for RFF speech stimuli for German speakers (n = 19) and age-matched English speakers (n = 19).

Target Consonant		Stimuli (VCV tokens in **bold** and underline)	Mean Token-Level RFF SD (ST) [95% CI]
/f/	Deutsch	In der **Auffah**rt wächst n**ie vie**l **Efe**u.	0.35 [0.31, 0.39]
		D**ie vie**r Personen gehen d**a vo**rne in d**ie Fa**brik.	0.44 [0.38, 0.49]
		D**ie vo**rgetragene **Aufführ**ung kommt b**ei fa**st allen Zuhörern gut an.	0.39 [0.32, 0.45]
	English	Only w**e fee**l you d**o fai**l in n**ew fa**llen dew.	0.44 [0.39, 0.49]
/ʃ/	Deutsch	D**ie Scha**r Kinder nimmt d**ie Sche**re mit in d**ie Schu**le.	0.37 [0.31, 0.42]
		M**aschi**nen können helfen, B**auschu**tt wegz**uschie**ben.	0.53 [0.40, 0.64]
		Sie sah ihn d**ie schö**nen, s**o schi**cken Teller z**uschie**ben.	0.33 [0.26, 0.38]
	English	The d**ew shi**mmered over m**y shi**ny bl**ue she**ll again.	0.69 [0.37, 0.90]
/f/	Deutsch	/**afa afa afa**/	0.25 [0.22, 0.28]
		/**ifi ifi ifi**/	0.22 [0.19, 0.24]
		/**ufu ufu ufu**/	0.22 [0.20, 0.23]
	English	/**afa afa afa**/	0.34 [0.28, 0.38]
		/**ifi ifi ifi**/	0.35 [0.29, 0.40]
		/**ufu ufu ufu**/	0.32 [0.28, 0.35]

Conclusions:

All the sentences except for “Maschinen können helfen, Bauschutt wegzuschieben” were similarly low in variability (< 0.5 ST). Since the variability is similar, “In der Auffahrt wächst nie viel Efeu” and “Die Schar Kinder nimmt die Schere mit in die Schule” are suggested to be used for German RFF because of their fluent readability and brevity. The novel German RFF sentences demonstrated lower token-level variability compared to the validated English RFF sentences. This indicates that the German stimuli have favorable stability for measuring vocal effort, providing confidence in their utilization for future RFF research. More broadly, this study establishes a systematic process for developing ecologically valid, continuous speech RFF stimuli in languages beyond English. The creation of linguistically diverse stimuli will enable cross-linguistic comparisons and expand the accessibility of vocal effort measurement through RFF for multilingual speakers.

Acknowledgements:

Research reported in this publication was supported by the National Institute On Deafness and Other Communication Disorders of the National Institutes of Health under Award Number R01DC018026. We thank Dr. Laura Fröhlich, head of Audiology of the department of Otorhinolaryngology—Head and Neck Surgery, University Hospital Bonn, Bonn, Germany, for making available a sound-treated clinic room for audio recordings.

References:

1. Kempster GB, Gerratt BR, Verdolini Abbott K, Barkmeier-Kraemer J, Hillman RE. Consensus auditory-perceptual evaluation of voice: Development of a standardized clinical protocol. *American Journal of Speech-Language Pathology*. 2009;18(2):124–32. DOI: 10.1044/1058-0360(2008/08-0017).

2. Lien Y-AS, Gattuccio CI, Stepp CE. Effects of phonetic context on relative fundamental frequency. *Journal of Speech, Language, and Hearing Research*. 2014;57(4):1259–67. DOI: 10.1044/2014_JSLHR-S-13-0158.

3. Gao Y, Dietrich M, DeSouza GN. Classification of vocal fatigue using sEMG: Data imbalance, normalization, and the role of Vocal Fatigue Index scores. *Applied Sciences*. 2021;11(10):4335. DOI: 10.3390/app11104335.

INFLUENCE OF TYPE I THYROPLASTY IMPLANT STIFFNESS AND LOCATION ON THE GLOTTAL PRE-PHONATORY SHAPE, STIFFNESS AND VIBRATION

Weili Jiang[1], Jacob Michaud-Dorko[2], Mahdi Sangbori[1], Liran Oren[3], Charles Farbos de Luzan[3], Ephraim Gutmark[4], Xudong Zheng[1], Qian Xue[1]

1 Department of Mechanical Engineering, Rochester Institute of Technology, Rochester, NY, USA
2 Department of Biomedical Engineering, University of Cincinnati, Cincinnati, OH, USA
3 Department of Otolaryngology-Head and Neck Surgery, University of Cincinnati, Cincinnati, OH, USA
4 Department of Aerospace Engineering, University of Cincinnati, Cincinnati, OH, USA

***Keywords*: Type I Thyroplasty Implant; Sound Production; Numerical Simulation**

Abstract:

Objectives / Introduction:

Thyroplasty Type I (TT1) is a common procedure for treating unilateral vocal fold paralysis. The procedure involves creating a window in the thyroid cartilage, allowing for the placement of an implant within the paraglottic space. This procedure aims to improve voice outcomes by narrowing the glottal gap between the paralyzed and healthy folds, thereby reducing leakage and threshold pressure. Traditionally, surgeons manually carve the implant, determining its specifications, including location, stiffness, and shape, through a trial-and-error approach. The impact of these implant characteristics on voice outcomes remains an ongoing research topic.

Clinical observations indicate that medialization in the infraglottal region yields better voice outcomes than medialization in the glottal region, although the exact reason remains unknown. In a study by Zhang and Chhetri [1], local medialization using a 2mm-diameter wood stick on excised human larynges showed improved glottal closure and higher-order harmonic excitation in voice when performed at a more inferior location. They hypothesized that inferior medialization contributed to a thicker vocal fold, but the medial surface's shape, including thickness, was not quantified, leaving the hypothesis unconfirmed. Zhang et al. [2] provided detailed structural changes before and after implant insertion using magnetic resonance imaging on ex vivo human larynges. They observed increased vertical thickness and significant stretching of the vocal fold tissue in the coronal plane, leading to a decrease in thyroarytenoid (TA) muscle dimension in the medial-lateral direction. However, since phonation experiments were not conducted, the direct relationship between structural changes and the produced sound is unclear.

The current study aims to investigate the impact of TT1 implant location and stiffness on vocal fold dynamics and voice outcomes using numerical methods, establishing a direct link among the implant characteristics, vocal fold prephonatory conditions, and sound characteristics. Compared to glottal-level medialization, we hypothesize that infraglottal medialization will create a slightly divergent pre-phonatory shape and a more substantial vertical stiffness gradient, potentially enhancing the produced sound.

Methods:

Geometry model. The geometric models of the larynx and implant were created from computed tomography scans of an ex-vivo canine larynx. Both the cartilage and vocal fold were reconstructed. The vocal fold was represented as a two-layer structure consisting of the cover and body layers.

Material property. Indentation tests were performed using a sphere indenter on the medial surface of two canine larynges (Figure 1). The tissue was indented with a maximum of 2 mm. The indentation was performed at three coronal planes from anterior to posterior sides and eleven vertical locations from the superior edge to 5 mm below the superior edge. In each location, a force-displacement relationship was obtained.

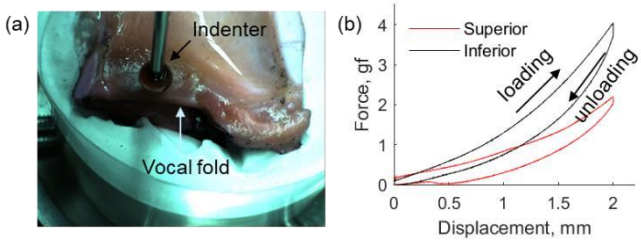


Figure 1. Indentation measurement. (a) Setup. (b) Example for force-displacement curves from inferior and superior measurements at the mid-coronal plane.

In the numerical simulation, the vocal fold was assumed to be a two-layer structure, including the body and cover layers. The material was assumed to be fiber-reinforced in each layer, and a vertical stiffness gradient exists in the cover layer [3], [4]. Material parameters were determined by matching the indentation measurement. The implant employed the Neo-Hookean material model with a stiffness ranging from 5 kPa to 1400 kPa.

Simulation setup. The simulation comprises two steps. In the first step, we insert the implant and perform numerical indentation with the inserted implant on the medial surface. The obtained stiffness distribution was used to compare with the experiment measurement. In the second step, we performed a fluid-structure interaction (FSI) simulation with the inserted implant to capture vocal fold oscillation, glottal flow waveform, and sound generation [4], [5]. The glottal flow employed a one-dimensional flow model, and the tissue deformation was modeled using a finite element method. The height of the implant was constant and set to half of the vertical height of the fold. The vertical insert location of the implant was varied between the superior and inferior aspects of the fold medial surface.

Results:

In the baseline case of the numerical simulation, we inserted an implant with a height matching the vertical height of the fold at the glottic level. The insertion depth was adjusted so that 80% of the medial surface was adducted to the midline. The implant stiffness was 120 kPa. The results show that the implant could effectively adduct the fold toward the midline with only a small leakage in the posterior side, but the shape of the medial surface profile does not show a significant difference before and after the implant was inserted (Figure 2a). The stiffness gradient was calculated between the superior and inferior sides of the medial surface along three Anterior-Posterior locations (Figure 2b). It shows that with the implant inserted, the stiffness gradient increased in all three locations and that the largest increase occurred on the anterior side.

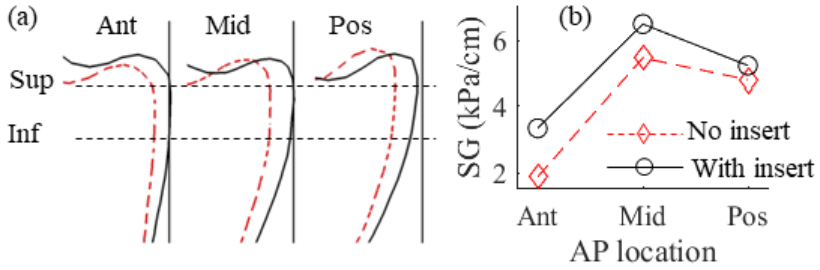


Figure 2. Influence of implant on pre-phonatory status from numerical simulations: (a) medial surface profile; (b) stiffness gradient (SG).

Conclusions:

The preliminary results show that the numerical model can simulate the impact of TT1 implant on vocal fold deformation and stiffness distribution before phonation. In subsequent simulations and analysis, we will change the implant height, vertical location, and stiffness. The effects of these parameters on vocal fold pre-phonatory deformation and stiffness will be investigated. Additionally, FSI simulations will be performed before and after the implant is inserted. The relationships among implant characteristics, vocal fold pre-phonatory conditions, vocal fold vibrations, and voice acoustics will be studied. We will discuss the results by comparing them with the experimental measurements [1,6,7].

Acknowledgements:

The research was funded by NIH Grant No. R01DC009435 from the National Institute on Deafness and Other Communication Disorders (NIDCD).

References:

1. Z. Zhang and D. K. Chhetri, “Effect of changes in medial surface shape on voice production in excised human larynges,” J Acoust Soc Am, vol. 146, no. 5, pp. EL412–EL417, 2019.
2. Z. Zhang, L. Wu, R. Gray, and D. K. Chhetri, “Three-dimensional vocal fold structural change due to implant insertion in medialization laryngoplasty,” PLoS One, vol. 15, no. 1, p. e0228464, Jan. 2020.
3. G. Dhondt, The Finite Element Method for Three-Dimensional Thermomechanical Applications. Wiley, 2004.
4. B. Geng, M. Movahhedi, Q. Xue, and X. Zheng, “Vocal fold vibration mode changes due to cricothyroid and thyroarytenoid muscle interaction in a three-dimensional model of the canine larynx,” J Acoust Soc Am, vol. 150, no. 2, pp. 1176–1187, 2021.
5. M. Movahhedi, B. Geng, Q. Xue, and X. Zheng, “A computational framework for patient-specific surgical planning of type 1 thyroplasty,” JASA Express Lett, vol. 1, no. 12, Dec. 2021.
6. A. Maddox, L. Oren, C. Farbos de Luzan, R. Howell, E. Gutmark, and S. Khosla, “An Ex-vivo Model Examining Acoustics and Aerodynamic Effects Following Medialization with and without Arytenoid Adduction,” Laryngoscope, vol. 133, no. 3, pp. 621–627, Mar. 2023.
7. L. Oren, A. Maddox, C. Farbos de Luzan, C. Xie, G. Dion, E. Gutmark, S. Khosla, “Acoustics and aerodynamic effects following glottal and infraglottal medialization in an excised larynx model,” European Archives of Oto-RhinoLaryngology. 2024. In print.

CAN THE VOCAL BEHAVIOR OF A MOUSE SERVE AS A MARKER FOR VOCAL FOLD PROPERTIES?

Tobias Riede[1]

[1] Department of Physiology, Midwestern University, Glendale, AZ, USA

***Keywords*: e.g. Voice; Mouse; Vocal fold injury; Vocal changes**

Abstract:

Objectives / Introduction: Currently, there is no model whose natural behavior can serve as a marker for the onset or the time course of a vocal fold pathology. Here I present preliminary data on the time-course of vocal changes after two experimental interventions affecting vocal fold properties in California mice.

California mice produce a call type ("SV call") to communicate over large distances. SV calls consist of 1 to 8 syllables, each approximately 250 ms in duration and with a fundamental frequency of 12 to 21 kHz. Light gas experiments demonstrated that SV calls are generated via airflow-induced self-sustained tissue/vocal fold vibration, i.e., fundamental frequency remains constant in the light gas atmosphere and critically depends on tissue viscoelastic properties (Riede et al. 2022). 3D reconstruction of laryngeal elements provides evidence that the mouse's vocal apparatus constitutes a highly nonlinear system.

This study investigated vocal changes after two types of interventions: (a) dehydration, and (b) vocal fold injury. Predictive vocal changes should occur if vocal fold properties are critical to the acoustic parameters of the mouse's vocal behavior.

Methods: *Vocal fold injury*: Unilateral epithelial injury was performed in male and female California mice to investigate whether injury and subsequent wound healing are associated with characteristic vocal changes. Six-month-old animals have undergone unilateral vocal fold injury (N = 3 per sex) or served as controls that received a sham operation (N = 2 per sex). Vocal activity was monitored continuously before and after the injury. Vocal changes were determined by comparing fundamental frequency, sound intensity, and the amount of nonlinear phenomena before and after injury. The size of the vocal fold injury was estimated in vocal fold coronal sections performed on excised tissue.

Dehydration: Water access was restricted in 8 California mice for 3 consecutive days to study the effect of acute systemic dehydration on vocal function. Each animal was continuously acoustically monitored for 24 hours before experimental onset, during the 3 days without water access, and for 2 days after the beginning of the rehydration. Body mass was measured before the experiment and at the end of day 3 without water access to estimate the degree of dehydration. SV syllables were characterized by fundamental frequency and sound intensity.

Results: We present results from two experiments intended to alter vocal fold properties.

Vocal fold injury: Only 2 of 6 mice did resume vocal behavior within 4 weeks after vocal fold injury. Their vocal activity was comparable to the activity pre-injury. Fundamental frequency and sound intensity have decreased after injury in both animals. More than 90% of syllables contained nonlinear phenomena which is an increase from 10% pre-injury. Histological sections indicated that the vocal fold injuries were smaller in the 2 mice that resumed vocal behavior after injury.

Dehydration: Both fundamental frequency and sound intensity decreased in 8 animals during the 3 days without water access. A partial recovery of fundamental frequency and sound intensity was observed after one day of rehydration.

Conclusions: Mild vocal fold injury and systemic dehydration caused vocal changes which confirms that vocal folds are critical for the acoustic properties of SV syllables produced by the mouse. The magnitude of the vocal fold injury seems to be associated with the likelihood of the mouse to resume vocal behavior. Vocal changes resemble those observed in humans.

Acknowledgements: This work was supported in part by NIH/NIDCD (R21DC019992).

References: Riede, T., Kobrina, A., Bone, L., Darwaiz, T. and Pasch, B., 2022. Mechanisms of sound production in deer mice (Peromyscus spp.). *Journal of Experimental Biology*, *225*(9), p.jeb243695.

LARYNGEAL AND UPPER AIRWAY SEQUELAE IN POST-ACUTE INFLUENZA INFECTION

Alexander G. Foote, PhD, CCC-SLP[1], Xin Sun, PhD[1]

[1] Department of Pediatrics, University of California – San Diego, La Jolla CA ,92092

***Keywords*: larynx, viral transmission, epithelial regeneration, single-cell analysis**

Abstract:

Objectives / Introduction: The cause for laryngeal and vocal fold sensorimotor vagal neuropathies is unclear.[1] Speculation into involvement of peripheral local inflammation and sensory neural cell changes following viral infection have been cited,[2] however, limited data exists to support such claims. Influenza A virus (IAV) is rapidly detected in the airways by epithelial and immune cells to induce antiviral inflammatory responses.[3, 4] The purpose of this study was to test the central hypothesis that IAV-mediated infection of the larynx and upper airway will activate epithelial and immune cells and remodel sensory afferent neural-circuits – contributing to mucosal inflammation and hypersensitivity.

Methods: Wildtype 8wk old mice were utilized to investigate epithelial tropism and immune response following IAV injury for repair strategies at 3-, 7- and 21-days post-infection (dpi). Whole upper airway tissue (hypopharynx/larynx/trachea) was collected at 7dpi for single-cell analysis. To investigate neural-circuit structural changes, we utilized a vesicular glutamate transporter (Vglut2) Cre line for sensory neuron-specific lineage tracing at 21dpi. Epithelial remodeling and immune response were assessed via histology, immunofluorescence and single-cell bioinformatic analysis, while neural-circuit innervating networks were quantified in whole-mount cleared 3D tissue.

Results: Viral mRNA transcripts to the upper airway of wildtype epithelium were most appreciated at 3dpi with viral clearance exhibited at 8dpi. Respiratory epithelium of the subglottis and trachea exhibited increased vulnerability to viral infection, with stratified squamous epithelia of the larynx and hypopharynx largely void of disease. Acute airway remodeling consisted of K5+K17+ basal cell hyperplasia, alongside, increased intraepithelial Ly6G+ neutrophil and CD3+ T cell infiltrates. Abberant epithelial and immune changes exhibited complete resolution at 21dpi, however, Vglut2-lineage traced sensory nerves displayed hyperinnervation following IAV compared to saline controls. Single-cell analysis revealed diversity of epithelial cells to the upper airway with specialized response to viral infection.

Conclusions: Our work utilized advanced histologic and bioinformatic approaches to explore a novel injury model of the larynx and upper airway. We demonstrate unique changes to structural cells and immune response that culminate in hyperinnervation of Vglut2+sensory neural circuits. Further work is necessary to demonstrate a causal link for viral-induced laryngeal pathophysiology and contribution for sensorimotor abnormalities.

Acknowledgements:

This work was funded by support from the NIH NIDCD 1R01AT011676-01 and the NIH NIDCD F32 DC021634-01.

References:

1. Sundar KM, Stark AC, Hu N, Barkmeier-Kraemer J. Is Laryngeal Hypersensitivity the Basis for Unexplained or Refractory Chronic Cough? ERJ Open Research. 2021.
2. Murry T, Branski RC, Yu K, Cukier-Blaj S, Duflo S, Aviv JE. Laryngeal sensory deficits in patients with chronic cough and paradoxical vocal fold movement disorder. Laryngoscope. 2010;120(8):1576-81. Epub 2010/06/22. doi: 10.1002/lary.20985. PubMed PMID: 20564660.
3. Barr J, Gentile ME, Lee S, Kotas ME, Fernanda de Mello Costa M, Holcomb NP, Jaquish A, Palashikar G, Soewignjo M, McDaniel M, Matsumoto I, Margolskee R, Von Moltke J, Cohen NA, Sun X, Vaughan AE. Injury-induced pulmonary tuft cells are heterogenous, arise independent of key Type 2 cytokines, and are dispensable for dysplastic repair. Elife. 2022;11. Epub 2022/09/09. doi: 10.7554/eLife.78074. PubMed PMID: 36073526.
4. Verzele NAJ, Chua BY, Law CW, Zhang A, Ritchie ME, Wightman O, Edwards IN, Hulme KD, Bloxham CJ, Bielefeldt-Ohmann H, Trewella MW, Moe AAK, Chew KY, Mazzone SB, Short KR, McGovern AE. The impact of influenza pulmonary infection and inflammation on vagal bronchopulmonary sensory neurons. Faseb J. 2021;35(3):e21320. Epub 2021/03/05. doi: 10.1096/fj.202001509R. PubMed PMID: 33660333.

EXPLORING THE EFFECTS OF PSYCHOSOCIAL STRESS ON THE LARYNGEAL MICROBIOME

Anumitha Venkatraman[1], John Binns[1], Katelyn Jacobs[3], Federico Rey[2], Seth Pollak[3], Susan Thibeault[1]

[1] Department of Surgery, Division of Otolaryngology- Head and Neck Surgery, University of Wisconsin Madison, Madison, WI, USA
[2] Department of Bacteriology, University of Wisconsin Madison, Madison, WI, USA
[3] Department of Psychology, University of Wisconsin Madison, Madison, WI, USA

***Keywords*: e.g. laryngeal microbiology; psychosocial stress; laryngeal biology,**

Abstract:

Introduction: Psychosocial stress and laryngeal physiology are circularly and inextricably linked.[1–3] Whereas 25% of individuals with voice problems report concurrent psychosocial stress,[3] a single exposure to stress can result in acoustic voice changes, and laryngeal muscle activation patterns.[4–7] However, little is known on the underlying biological mechanisms of psychosocial stress, in laryngeal physiology. The microbiome is a key mediator of the psychosocial stress response in other organs such as the gut, vagina and brain, resulting in reduced microbial diversity and abundance.[8–10] Laryngeal microbiology is an emerging area of research[11–13] that may shed light on the elusive role of psychosocial stress in laryngeal physiology.

Objectives: To delineate the effects of two psychosocial stress protocols (acute and chronic) on laryngeal microbiota. We hypothesize that both acute and chronic psychosocial stress will result in reduced microbial diversity and abundance in the larynges of stressed mice when compared to control animals. Chronic exposure to stress results in more severe microbial responses in other organs (e.g. gut).[14,15] In addition, females have differential responses to stress in the gut microbiome due to natural estrogenic variation,[16] with dysphonia and mental health conditions being more prevalent in this biological sex. Thus, we hypothesize that chronic stress exposure and females (across stress paradigms) may result in more pronounced effects in the laryngeal microbiota.

Methods: Sixty C57BL/6 mice (6-10 weeks of age, 30 F, 30 M) were randomly divided into acute stress, chronic stress, and control groups (10M, 10F per group). Animals in the acute stressed group experienced restraint stress for 6 hours a day for 7 days, whereas those in the chronic stress group experienced the same for 6 hours a day for 14 days. Corticosterone assays were performed on plasma samples collected from blood before and following each protocol to confirm stress induction. Following each psychosocial stress protocol, bacterial DNA were extracted from the larynx and amplified for 16S RNA gene sequencing. QIIME2 and R studio were used for sequence data bioinformatic analysis.

Results: Chronic – but not acute – psychosocial stress resulted in a significant shift in laryngeal microbiota composition. Within the differentially abundant taxa of the chronic stress group, there was an increased abundance of *Streptococcus* and a decreased abundance of *Corynbacterium*. Whereas increased abundance of *Streptococcus* has been reported in certain laryngeal pathologies,[12,13] *Corynbacterium* may regulate allergen sensitization in other organs.[17] With respect to sex-related differences, females experiencing chronic stress had more pronounced microbial differences in measures of diversity (reduced alpha diversity) and abundance of certain bacterial taxa.

Conclusions: Psychosocial stress exposure has a significant effect on laryngeal microbial composition, however, the duration of the stressor (chronic versus acute) matters. In addition, females were more susceptible to stress-induced laryngeal microbiota variation. These data lay the groundwork for delineating the functional consequences of stress-induced phylogenic differences on laryngeal physiology.

Acknowledgements:

The authors would like to thank Ran An, Stephanie Bartley, Vlasta Lungova and Se in Kim for their help and contributions on aspects of data collection protocols and analysis.

References:

1. Baker J, Ben-Tovim D, Butcher A, Esterman A, McLaughlin K. Psychosocial risk factors which may differentiate between women with Functional Voice Disorder, Organic Voice Disorder and a Control group. *International Journal of Speech-Language Pathology*. 2013;15(6):547-563. doi:10.3109/17549507.2012.721397

2. Nguyen-Feng VN, Asplund A, Frazier PA, Misono S. Association Between Communicative Participation and Psychosocial Factors in Patients With Voice Disorders. *JAMA Otolaryngol Head Neck Surg*. Published online December 23, 2020. doi:10.1001/jamaoto.2020.4956

3. Dietrich M, Verdolini Abbott K, Gartner-Schmidt J, Rosen CA. The Frequency of Perceived Stress, Anxiety, and Depression in Patients with Common Pathologies Affecting Voice. *Journal of Voice*. 2008;22(4):472-488. doi:10.1016/j.jvoice.2006.08.007

4. Perrine BL, Scherer RC. Aerodynamic and Acoustic Voice Measures Before and After an Acute Public Speaking Stressor. *J Speech Lang Hear Res*. 2020;63(10):3311-3325. doi:10.1044/2020_JSLHR-19-00252

5. Pisanski K, Nowak J, Sorokowski P. Individual differences in cortisol stress response predict increases in voice pitch during exam stress. *Physiology & Behavior*. 2016;163:234-238. doi:10.1016/j.physbeh.2016.05.018

6. Helou LB, Rosen CA, Wang W, Verdolini Abbott K. Intrinsic Laryngeal Muscle Response to a Public Speech Preparation Stressor. *Journal of Speech, Language, and Hearing Research*. 2018;61(7):1525-1543. doi:10.1044/2018_JSLHR-S-17-0153

7. Dietrich M, Verdolini Abbott K. Psychobiological Stress Reactivity and Personality in Persons With High and Low Stressor-Induced Extralaryngeal Reactivity. *J Speech Lang Hear Res*. 2014;57(6):2076-2089. doi:10.1044/2014_JSLHR-S-12-0386

8. Foster JA, Rinaman L, Cryan JF. Stress & the gut-brain axis: Regulation by the microbiome. *Neurobiol Stress*. 2017;7:124-136. doi:10.1016/j.ynstr.2017.03.001

9. Wu WL, Adame MD, Liou CW, et al. Microbiota regulate social behaviour via stress response neurons in the brain. *Nature*. 2021;595(7867):409-414. doi:10.1038/s41586-021-03669-y

10. Gur TL, Bailey MT. Effects of Stress on Commensal Microbes and Immune System Activity. *Adv Exp Med Biol*. 2016;874:289-300. doi:10.1007/978-3-319-20215-0_14

11. An R, Gowda M, Rey FE, Thibeault SL. Selective Bacterial Colonization of the Murine Larynx in a Gnotobiotic Model. *Front Microbiol*. 2020;11:594617. doi:10.3389/fmicb.2020.594617

12. Hanshew A, Jetté M, Thibeault S. Characterization and comparison of bacterial communities in benign vocal fold lesions. *Microbiome*. 2014;2:43.

13. Jetté M, Dill-McFarland K, Hanshew A, Suen G, Thibeault S. The human laryngeal microbiome: effects of cigarette smoke and reflux. *Scientific reports*. 2016;6.

14. Kelly LS, Apple CG, Gharaibeh R, et al. Stress-related changes in the gut microbiome after trauma. *J Trauma Acute Care Surg*. 2021;91(1):192-199. doi:10.1097/TA.0000000000003209

15. 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

16. Pigrau M, Rodiño-Janeiro BK, Casado-Bedmar M, et al. The joint power of sex and stress to modulate brain–gut–microbiota axis and intestinal barrier homeostasis: implications for irritable bowel syndrome. *Neurogastroenterology Motil*. 2016;28(4):463-486. doi:10.1111/nmo.12717

17. Chun Y, Do A, Grishina G, et al. The nasal microbiome, nasal transcriptome, and pet sensitization. *Journal of Allergy and Clinical Immunology*. 2021;148(1):244-249.e4. doi:10.1016/j.jaci.2021.01.031

EXPLORING THE MOLECULAR BASIS OF REINKE'S EDEMA

Markus Gugatschka, Tanja Grossmann, Magdalena Grill

Division of Phoniatrics, Medical University Graz, Austria

Keywords: Reinke's edema; molecular laryngology; vocal fold biology

Abstract:

Objectives / Introduction: The diagnosis of Reinke's edema (RE) is a classic visual diagnosis. This and surgical therapy have pushed questions about the pathogenesis into the background. Smoking, vocal strain and hormonal influences are classically cited as causes, but the exact effect on the vocal fold tissue is not known. The aim of this paper is to provide an overview of the molecular and micro-anatomical basis of RE.

Methods: Literature review, summary of own experiments.

Results: The understanding of molecular mechanisms and electron-microscopical findings led to a better understanding of the pathophysiology. Our own in vitro experiments showed that the combination of cigarette smoke and vibration promotes neo-angiogenesis. Elongated and ectatic vessels are typical laryngoscopic findings of RE. The permanent vibration of these elongated microvessels within the vocal folds leads to leakage with increased extravasation (Sato et al.) and the resulting edema. However, the vocal fold fibroblasts themselves also change as a result of years of exposure to cigarette smoke.

Conclusions: The combination of recent in vitro experiments and known electron microscopic findings provide a clearer picture of the pathogenesis of RE. In vitro experiments allow for the first time the separate and combined consideration of risk factors (smoking, vibration). A causal understanding of the pathophysiology creates the basis for therapies that are not purely symptomatic. The use of various lasers for the treatment of RE is based precisely on this understanding.

References:

Electron microscopic and immunohistochemical investigation of Reinke's edema. Sato K, Hirano M, Nakashima T. Ann Otol Rhinol Laryngol 1999; 108:1068–1072.

Exploring the Pathophysiology of Reinke's Edema: The Cellular Impact of Cigarette Smoke and Vibration. Grossmann T, Steffan B, Kirsch A, Grill M, Gerstenberger C, Gugatschka M. Laryngoscope. 2021 Feb;131(2):E547-E554

Vocal Fold Fibroblasts in Reinke's Edema Show Alterations Involved in Extracellular Matrix Production, Cytokine Response and Cell Cycle Control. Grill M, Lazzeri I, Kirsch A, Steurer N, Grossmann T, Karbiener M, Heitzer E, Gugatschka M. Biomedicines. 2021 Jun 26;9(7):735.

DIFFERENTIAL CYTOKINE RESPONSE AND ALTERED DNA METHYLATION IN VOCAL FOLD FIBROBLASTS OF REINKE'S EDEMA PATIENTS

Annamaria Fricska[1], Nina Steurer[1, 2], Tanja Grossmann[1], Barbara Steffan[1], Ellen Heitzer[3], Markus Gugatschka[1], Magdalena Grill[1]

[1] Division of Phoniatrics, Department of Otorhinolaryngology, Medical University of Graz, Graz, Austria

[2] current affiliation: Evotec SE, Hamburg, Germany

[3] Institute of Human Genetics, Diagnostic & Research Center for Molecular BioMedicine, Medical University of Graz,Graz, Austria.

***Keywords*: Reinke's edema, vocal fold fibroblasts, cytokines, DNA methylation analysis**

Abstract:

Objectives / Introduction:

Reinke's edema (RE) is a benign lesion of the vocal folds (VF) mostly occurring in middle-aged women, leading to a hoarse, non-sustainable and low-pitched voice and sometimes breathing difficulties. Current treatment options are phonosurgery and cessation of risk factors such as smoking and heavy voice use. But even after treatment, a full recovery cannot be regained and there is an urgent need for a better understanding of the molecular causes of the disease and new treatment options.

Previously published RNA sequencing (RNAseq) data in our lab (Grill et al. 2021) led to the hypothesis that human vocal fold fibroblasts (hVFF) from RE patients have an altered response to cytokines, and are permanently altered by differential methylation.

Methods:

To test this, hVFF from our cell bank, previously collected from female RE patients and controls during phonosurgery or post-mortem (Diagnostic and Research Institute of Pathology, Medical University of Graz), were cultured in DMEM containing 10% FBS and 100 µg/ mL Normocin.

Cytokine treatment: Primary hVFF from RE and controls were starved over night with DMEM, 0.5% FBS and 100 µg/ mL Normocin. Cells were treated with or without either 10 ng/ mL TGFβ, 5 ng/ mL IL-1β, or 25 ng/ mL IL-6 in starving medium, respectively, for 15 min, 1 h, 6 h or 24 h (TGFβ: n = 6, IL-6: n = 3, IL-1β: n = 3). RNA levels were measured by qPCR, protein levels by western blot analysis.

Untreated cells were analyzed for DNA methylation (control n = 11, RE n = 17). Therefore, cell pellets were lysed and genomic DNA was isolated, purified and bisulfite converted, and DNA methylation was measured by Illumina DNA methylation EPIC v2 array. Bioinformatical analysis was done using the R packages minfi, shinyEPICo and seSAMe.

Histology: For evaluation of tissue morphology (control n = 2, RE n = 4), 5 µm paraffin-embedded tissue sections of control and RE VF tissue received from the department of pathology, were stained with haematoxylin and eosin according to standard procedures. Further, immunohistochemistry (IHC) was used to determine celltype-specific localization of the targets of interest, such as AQP1, TGM2 and IL-1R1.

Results:

TGFβ treatments: Treatment of hVFF with TGFβ resulted in an altered, increased gene expression of HIF1α, AQP1 and TGM2 in RE hVFF compared to control hVFF.

IL-6 and IL-1β treatments: Treatment of primary hVFF resulted in a cellular response to both, IL-1β as well as IL-6, shown by phosphorylation of the transcription factors NFκB and STAT3, respectively. While previous RNAseq experiments revealed an increased expression of IL-1R1 and IL-6R in RE hVFF compared to controls, we detected a further increase of IL-1R1 mRNA expression upon 24h of IL-1β or IL-6 treatments in hVFF from RE patients, but not in healthy controls or untreated RE hVFF.

DNA methylation: preliminary data analysis: non-directed analysis of differential methylation positions (DMP) revealed 7 hypermethylated and 17 hypomethylated positions (CpG sites) with an FDR < 0.05, three of them positioned in the shore, one in shelf and all others in OpenSea positions. Further, targeted analysis of DMP was done for differentially expressed genes previously determined by RNAseq.

IHC: While in control tissue, AQP1 was mainly expressed in endothelial cells and perivascular, AQP1 in RE VF was found highly expressed throughout the lamina propria. TGM2 in control tissue was mainly expressed in endothelial cells and perivascular, whereas TGM2 in RE VF was also expressed in fibroblasts throughout the lamina propria. In control tissue, IL-1R1 was mainly detected in epithelial, endothelial, and some single, isolated cells, but IL-1R1 staining in RE VF was found throughout the lamina propria. Fibroblasts, endothelial cells and other vimentin-positive cells were evenly stained in control and RE VF tissue, confirming an equal stainability of both, control and RE, tissue samples and excluding tissue preparation/ fixation issues as a cause for differential staining.

Conclusions:

We were able to show that hVFF isolated from RE patients do indeed have an altered response to treatment with the cytokines TGFβ, IL-1β or IL-6, respectively, leading to increased gene expression levels of HIF1α, AQP1, TGM2 and IL-1R1. In our preliminary DNA methylation data analysis, we found differentially methylated positions in RE hVFF compared to control hVFF that have to be confirmed in follow-up experiments. In preliminary IHC analysis of FFPE tissue, we were able to show that an altered gene expression of some targets of interest is also detected in RE tissue sections. Based on these results, we suggest epigenetic modifications that lead to an altered regulation of gene expression and contributes to the pathogenesis of RE.

Acknowledgements:

Thanks to L. Glawitsch for performing the EPIC v2 workflow and S. Hasenleithner for providing help with DNA methylation analysis using R. This work is supported by the Austrian Science Fund (FWF) grant P36067, awarded to M. Grill.

References:

Grill, M.; Lazzeri, I.; Kirsch, A.; et al. Vocal fold fibroblasts in Reinke's edema show alterations involved in extracellular matrix production, cytokine response and cell cycle control. Biomedicines. 2021;9(7)

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

Francisco Contreras-Ruston[1,2,3], Antoni Callen[4], Matias Zañartu[5], Castillo-Allende[6], Saavedra-Garrido, Jorge[7], Ochoa-Muñoz Andrés[7], Eric J. Hunter[8] , Jordi Navarra[1], Sonja A. Kotz[2]

[1] Department of Cognition, Development and Educational Psychology, University of Barcelona, Barcelona, Spain
[2] Faculty of Psychology and Neuroscience, Department of Neuropsychology & Psychopharmacology, Maastricht University, Maastricht, The Netherlands.
[3] Speech-Language Pathology and Audiology Department – Universidad de Valparaíso, San Felipe, Chile
[4] Institut de Recerca, Hospital Sant Joan de Déu, Barcelona, Spain
[5]Universidad Técnica Federíco Santa María, Viña del Mar, Chile
[6]Department of Communicative Sciences and Disorders, Michigan State University, EEUU
[7]Department of Statistics, Universidad de Valparaíso, Valparaíso, Chile
- [8]Department of Communication Sciences and Disorders, University of Iowa, Iowa City, EEUU

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

Abstract:

Objectives / Introduction:

Parkinson's disease (PD) often presents voice-specific symptoms such as a decrease in voice loudness. This potential dysregulation might partially result from an alteration in the feedforward system underlying (Guenther et al.,2006; Huang et al., 2019), further indicated that sensory feedback regulations rely on a forward system that supports voice regulation. We know little about how this system is affected in Individuals with PD (IwPD), and why they often fail to notice their difficulties in voice regulation. Self-awareness of one's own voice is a crucial element for voice clinics, where clinicians frequently use voice self-assessment instruments in PD's voice evaluation. However, likely due to their diminished self-voice awareness, PD may consider their voice regulation as normal in these assessments. Other approaches, such as electrophysiological studies are challenging but have shown to be potentially useful to test forward models by looking into the motor preparatory phase before voice onset and the outcome of sensory feedback with event-related brain potentials (ERPs). For example, Knolle et al., (2019) showed a motor induced suppression (MIS) effect of the N1 and P2 ERP components, that is sensitive to expected sensory feedback. Changes in MIS in IwPD might therefore be a good measure to monitor difficulties in voice loudness control (Emmendorfer et al., 2021; Abur et al., 2018; Li et al., 2021).

Methods:

To gain insight into the regulation of voice loudness control in IwPD, we conducted the two studies.

Study 1 - Questionnaire Study
IwPD and individuals with General Voice Disorders (GVD) completed the Voice Symptom Scale (VoiSS) questionnaire to evaluate their voice self-awareness. Vocal loudness (dB) was also assessed. Univariate and multivariate analyses were used to compare outcome measures between the two groups.

Study 2 - EEG Study
This study examined the sensory feedback of the patiens' own in terms of voice loudness in ERPs (N1 and P2) in IwPD. A Button-Press Paradigm (BP) with three conditions was used: (a) A Voice Motor Condition (VMC) x (+15dB/habitual loudness) required participants to press a button to induce their voice; (b) Voice Only Condition (VOC) x (+15dB/habitual loudness) where participants simply listened and attended to their pre-recorded voice; and (c) Motor-Only Condition (MOC) where participants pressed a button only. The experiment included 100 trials in each condition. In healthy individuals and in IwPD.

Results:

Study 1
IwPD reported significantly fewer vocal symptoms than those with GVD in all VoiSS questionnaire domains. Multivariate principal component analyses revealed no significant correlations between VoiSS scores and the voice acoustic parameters of the participants. Despite experiencing hypophonia, IwPD scored lower on the questionnaire within the healthy voice range. Hierarchical Clustering Analysis grouped participants into three distinct categories, primarily based on age, vocal loudness, and VoiSS domain scores, distinguishing between PD and GVD individuals.

Study 2
The amplitudes of ERP components previously associated with suppression were found to be anomalous IwPD. This was evident both at the normal loudness and at +15dB levels, which differed in healthy control individuals but not IwPD.

Conclusions:

These studies highlight key differences in their own voice perception and production in IwPD. Study 1, showed that IwPD report fewer vocal symptoms than persons with other voice disorders, indicating a discrepancy in self-awareness of changes in voice regulation in PD. Study 2 revealed alterations in sensory feedback in PD, as seen in N1 amplitude changes, suggesting impaired neural processing in 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). Audi-tory 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: 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

EARLY ALZHEIMER'S DISEASE-RELATED LARYNGEAL INFLAMMATION AND MUSCLE PATHOLOGY IN THE TgF344-AD RAT MODEL

Denis Michael Rudisch[1,2,3], Maryann N. Krasko[1,2], Michelle R. Ciucci[1,2,4]

[1] Department of Communication Sciences & Disorders, University of Wisconsin-Madison, Madison, Wisconsin, USA
[2] Department of Surgery – Division of Otolaryngology, University of Wisconsin-Madison, Madison, Wisconsin, USA
[3] Institute for Clinical and Translational Research, University of Wisconsin-Madison, Madison, Wisconsin, USA
[4] Neuroscience Training Program, University of Wisconsin-Madison, Madison, Wisconsin, USA

Keywords*: *Alzheimer's Disease; Inflammation; TgF344-AD; Voice

Objectives / Introduction:

Alzheimer's Disease (AD) is a progressive neurodegenerative disease caused by abnormal presence of tau protein, beta-amyloid aggregation, and inflammation in the central and peripheral nervous systems. Although mostly known for hallmark signs such as memory loss and cognitive decline (dementia), communication deficits also occur during disease progression and significantly impact well-being and quality of life. However, little is known about peripheral changes, specifically laryngeal pathology in the prodrome and early stages of the disease, due to delayed clinical diagnosis. To overcome this obstacle, we use the validated transgenic *TgF344-AD* rat model of Alzheimer's disease that manifests *early* behavioral, cognitive, and neuropathological dysfunction. Our previous work with this model demonstrated ultrasonic vocalization deficits at 6mo of age and AD-related pathology and inflammation in the thyroarytenoid muscle (TA) at 12mo of age. The objective of this work is to test the central hypothesis that AD-pathology and inflammation manifest *early* in *several* laryngeal muscles that contribute to vocalization. We aim to test the hypotheses that 1) ADrelated pathology can be found in laryngeal tissue at 12mo of age (early stage) in the TA, cricothyroid (CT), and posterior cricoarytenoid (PCA) muscles, 2) AD-related inflammatory processes are present in the TA, CT, and PCA muscles in the early stages of disease, and to 3) establish that the *TgF344-AD* rat model of Alzheimer's disease is suitable for preclinical investigations of early AD-related laryngeal pathology.

Methods:

RT-qPCR was used to characterize peripheral inflammation and AD-related pathology via gene expressions in the TA, CT, and PCA muscles of six male *TgF344-AD* rats (n=6) and six WT Fischer-344 rats (n=6) at 12 months of age. Due to the ongoing controversy on aerodynamics and motor control of ultrasonic vocalizations, as well as the absence of the alar muscle in humans, TA, CT, and PCA were chosen, as they play a significant role in the production of vocalizations. One hemilarynx was used from each rat to assay for AD-specific peripheral gene expressions *Serpina3n, Uqcrc2, Pkp4, Ctsl, Bace2, Prkc ,* and *Igf2*; pro-inflammatory cytokines *Il1r1, Rorc, Jun, Foxp3, Myd88, Il1a, Il1β, Il6, Il12b, Il2, Ifnγ,* and anti-inflammatory markers *Il10*, and *Ifnα1* in the TA. PCR reactions were run in triplicate and normalized to *Hsp90* housekeeping gene. The same gene expression targets were used to assay CT and PCA muscles from the second half of the bisected larynges.

Results:

Our recent findings demonstrated significant downregulation of AD-related genes in *TgF344-AD* rats compared to WT Fischer-344 controls in the early stage of the disease (12mo) in *Uqcrc2* (FC = 0.208; p = 0.003; d = 6.712), *Bace2* (FC = 0.022; p = 0.019; d = 3.117), *Serpina3n* (FC = 0.017; p = 0.038; d = 2.502), and *Igf2* (FC = 0.461; p = 0.054; d = 2.200), as well as significant downregulation of pro-inflammatory gene *Myd88* (FC = 0.091; p = 0.008; d = 2.399). Based on these results, additional investigations are ongoing, and results will be presented for AD-related inflammatory markers *Il1r1, Rorc, Jun, Foxp3, Myd88, Il1a, Il1β, Il6, Il12b, Il2, Ifnγ, Il10*, *Ifnα1*, and AD-specific markers Serpina3n, Uqcrc2, Pkp4, *Ctsl, Bace2, Prkc* , and *Igf2* in the CT and PCA.

Conclusions:

Our study demonstrated significant *early-stage* downregulation of AD-related genes in the TA muscle in 12-month-old *TgF344-AD* rats that possibly contribute to related vocal deficits and indicated inflammatory dysregulation at the early stage of the disease. These results support current findings on early peripheral dysregulation and are the first to demonstrate early AD-related pathology in laryngeal muscles for vocalization. Further characterization of inflammatory processes and dysregulation in laryngeal tissue will help in understanding potential dichotomous activities of cytokines and inflammation cascades during the early disease stage. Preclinical investigations of laryngeal muscle pathology are imperative for improving our understanding for onset and progression of AD-related vocal dysfunction and overall AD pathophysiology for early detection and targeted interventions.

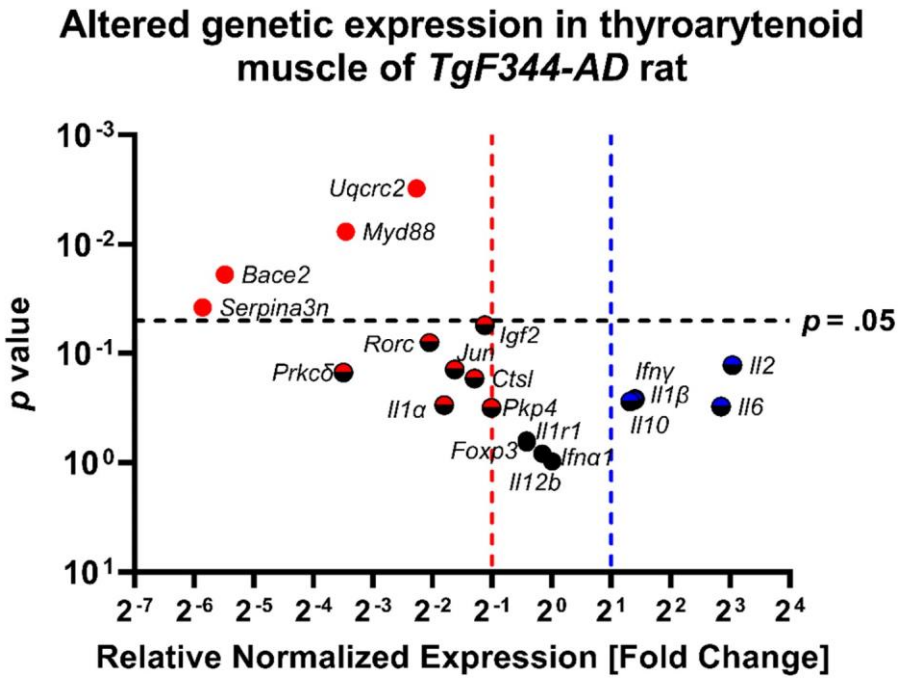
Acknowledgements:

This work was made possible through collaboration with Dr. Nadine P. Connor and the Connor Laboratory. Funding was received from the National Institute on Deafness and Other Communication Disorders, and the National Institute on Aging of the National Institutes of Health, R01AG085564, R01DC018071, R01DC018584, R01DC014358, R01AG082052, and T32DC009401.

References:

Rudisch DM, Krasko MN, Barnett DGS, Mueller KD, Russell JA, Connor NP and Ciucci MR (2024) Early ultrasonic vocalization deficits and related thyroarytenoid muscle pathology in the transgenic TgF344-AD rat model of Alzheimer's disease. *Front. Behav. Neurosci.* 17:1294648. doi: 10.3389/fnbeh.2023.1294648

Figure 1:



Our previous study demonstrated significantly downregulated genes associated with AD and inflammation (Rudisch et al., 2024). Each circle represents a specific gene target: Black indicates no significance; blue indicates upregulation; red indicates downregulation. Points to the right of the blue dashed line show upregulation, ≥ 2-fold. Points to the left of the red dashed line indicate downregulation. Points above the horizontal black dashed line indicates *p*-value ≤0.05, denoted with a solid-colored icon.

PRODROMAL VOCAL CHANGES IN HOMOZYGOUS PINK1 KNOCKOUT MODEL OF PARKINSON DISEASE

Maryann N. Krasko[1,2], Denis Michael Rudisch[1,2,3], Michelle R. Ciucci[1,2,4]

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

[2] Department of Surgery, Division of Otolaryngology-Head and Neck Surgery, University of Wisconsin-Madison, Madison, WI, USA

[3] Institute of Clinical and Translational Research, University of Wisconsin-Madison, Madison, WI, USA

[4] Neuroscience Training Program, University of Wisconsin-Madison, Madison, WI, USA

***Keywords*: Voice; Ultrasonic Vocalizations; Parkinson disease**

Abstract:

Objectives / Introduction:

Vocal deficits are some of the earliest and most common signs of Parkinson disease (PD);[1,2] yet, they are not identified or treated until mid-to-late stages of the disease. This is largely attributed to the fact that vocal changes begin to manifest in in the prodromal stage, prior to nigrostriatal dopamine depletion and the appearance of hallmark motor signs - and therefore formal diagnosis.[2] Moreover, the pathology underlying this vocal decline is currently unknown. This renders the identification and treatment of PD-associated vocal dysfunction in the prodromal stage nearly impossible. To overcome this limitation, we use the validated early-onset genetic *Pink1* knockout (-/-) rat. Previous work investigating vocal calls of *Pink1-/-* rats across the disease process revealed progressive decline in several vocal parameters, including intensity and bandwidth, as well as insoluble alpha-synuclein in the brainstem at 8 months of age.[3] Here, we aimed to assess vocalization at 4 months of age, analogous to the prodromal stage of disease, and quantify pathologic alpha-synuclein in the brainstem. We hypothesized that at 4 months of age *Pink1-/-* rats would show early changes to vocalizations, as well as presence of pathologic alpha-synuclein in brainstem nuclei associated with vocal function.

Methods:

Six *Pink1-/-* male rats and 8 male wildtype (WT) control rats were assessed in this pilot study. All rats were tested at 4 months of age, analogous to the prodromal stage of PD. Calls (ultrasonic vocalizations, USVs) were elicited via mating paradigm, after which vocalizations were recorded for 90 seconds. DeepSqueak software was used to categorize calls into simple or complex calls, and to extract the following outcomes: average USV principal frequency, peak frequency, low frequency, high frequency, bandwidth, intensity, call length, and call count. Following behavioral assessment, neural tissue was harvested and processed for immunohistochemistry. Proteinase-K-resistant alpha-synuclein will be examined in the nucleus ambiguus and nucleus tractus solitarius.

Results:

Findings revealed that *Pink1-/-* rat complex calls were produced with a significantly greater bandwidth (p = 0.035) and average high frequency (p = 0.034) compared to WT controls. *Pink1-/-* rat complex calls were also produced with greater mean power (p = 0.035) than that of WT complex calls. Alpha-synuclein immunohistochemistry in the brainstem is ongoing.

Conclusions:

Findings from this study reveal that the genetic knockout of the *Pink1-/-* gene results in early changes to vocalizations. Moreover, this will be the first study to assess proteinase-k-resistant alpha-synuclein in the nucleus ambiguus and the nucleus tractus solitarius in *Pink1-/-* rats at 4 months of age, which may provide insight into the pathology underlying these early vocal changes.

Acknowledgements:

This work was funded by the National Institute on Deafness and Other Communication Disorders (T32DC009401, R01DC018584).

References:

1. Ho, A.K.; Iansek, R.; Marigliani, C.; Bradshaw, J.L.; Gates, S. Speech impairment in a large sample of patients with Parkinson's disease. Behav. Neurol. 1998, 11, 131–137, doi:10.1155/1999/327643.
2. Ma, A., Lau, K. K., & Thyagarajan, D. (2020). Voice changes in Parkinson's disease: What are they telling us?. *Journal of Clinical Neuroscience*, *72*, 1-7.
3. Grant, L. M., Kelm-Nelson, C. A., Hilby, B. L., Blue, K. V., Paul Rajamanickam, E. S., Pultorak, J. D., ... & Ciucci, M. R. (2015). Evidence for early and progressive ultrasonic vocalization and oromotor deficits in a PINK1 gene knockout rat model of Parkinson's disease. Journal of neuroscience research, 93(11), 1713-1727.

IDEAL GLOTTAL WAVEFORM: BASIC CONCEPTS AND KINEMATIC MODELLING WITH SYNTHETIC KYMOGRAMS

Jan G. Švec[1], S. Pravin Kumar[2], O. Vencálek[3], Sarah Lehoux[1,4]

[1] Voice Research Lab, Dept. Experimental Physics, Faculty of Science, Palacký University, Olomouc, Czechia

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

[3] Department of Mathematical Analysis and Applications of Mathematics, Faculty of Science, Palacký University, Olomouc, Czechia

[4] Department of Head and Neck Surgery, University of California, Los Angeles, California, USA

***Keywords*: Glottal waveform; Kymography; Mucosal waves; Vocal fold oscillations**

Abstract:

Objectives / Introduction:

Wider implementation of high-speed laryngeal videoendoscopic and kymographic imaging into clinical practice requires a better understanding of the nature of the vocal fold oscillatory waveforms observed laryngoscopically. The motion of the upper margin of the vocal folds is normally delayed behind the motion of the lower margin, making the glottis convergent during the opening phase and divergent, or less convergent, during the closing phase. Detection of the vertical phase differences (VPDs) between the lower and upper margins is diagnostically important as it reveals about mucosal pliability. Proper detection of VPDs through laryngoscopy, however, requires the lower vocal fold margin to be visible during the glottal closing phase, which is not always the case in clinical practice. In this paper, we therefore define the "ideal glottal waveform" in which the glottis is convergent during opening and divergent during the closing phase, creating the most favorable conditions for measuring VPDs. In such a case, the glottal edge is determined by the movements of the upper vocal fold margins during the opening phase, and by the movements of the lower vocal fold margins during the closing phase. The aim of this study was to define the ideal glottal waveform quantitatively, and to identify the favorable kinematic conditions in which it occurs.

Methods:

We simulated different glottal waveforms using a previously developed two-dimensional mucosal-wave-based kinematic model of the vocal fold oscillations, which was found suitable for generating kymograms resembling those observed laryngoscopically in vivo [1, 2]. For this introductory study, the original curved shape of the M5 model [3] was made rectangular by setting the entrance and exit radii to be below 10 micrometers; this allowed the upper and lower margins to be uniquely determined as distinct points rather than curved surfaces and made the detection of the ideal waveforms easier. The vocal fold surface points were driven to oscillate sinusoidally in horizontal and vertical directions following circular trajectories in the transversal plane. We varied five model input parameters within clinically relevant ranges: amplitude of oscillation of the vocal fold's upper margin A_U (0.1 to 1.1 mm), amplitude of oscillation of the vocal fold's lower margin A_L (0.1 to 1.1 mm), VPDs between the upper and lower vocal fold margins φ_{VPD} (0° to 125°), glottal halfwidth H_W (-0.05 to 1.2 mm), and glottal convergence angle ψ_{CVG} (-15° to 35°). Each of these five parameters was set to five different values covering the considered ranges, resulting in a total of 3125 (i.e., 5^5) oscillatory waveforms. The rest of the input parameters were kept constant. For simplicity, the kinematic and geometric properties were set to be identical on the left and right vocal folds so that the resulting vibratory pattern was symmetric.

To quantify the degree of "ideality" of the glottal waveforms, we introduced "upper margin quotient, Q_U" quantifying the proportion of time when the upper margin is at glottal edge during opening, and "lower margin quotient, Q_L" quantifying the proportion of time when the lower margin is at glottal edge during closing. The ideal glottal waveform was defined as the case when $Q_U = Q_L = 1$.

Results:

Within our dataset of 3125 cases, the opening phase of the glottal waveform was entirely defined by the movement of the VF upper margin in 68 % (2134) of the cases (Q_U =1), whereas the closing phase was wholly defined by the movement of the lower margin in 28% (881) of the cases (Q_L=1). The ideal waveform, fulfilling both of these conditions simultaneously, occurred in 12 % of our dataset cases (368 out of 3125 cases).

The ideal waveforms occurred most likely when the upper and lower margin amplitudes A_U and A_L were similar, the glottal convergence angle ψ_{CVG} was close to zero, the glottal halfwidth H_W was below 0.45 mm, and the vertical phase differences φ_{VPD} were larger than 50° (Figure 1).

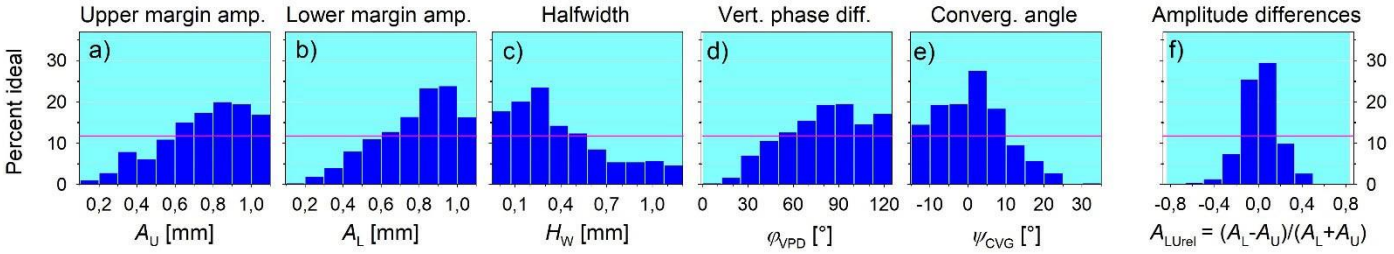


Figure 1: Percentages of ideal waveforms within different subranges of input parameter values. The pink horizontal lines indicate the average percentages of ideal waveforms for the whole dataset.

Discussion and conclusions:

Among the five basic input parameters, we found the convergence angle ψ_{CVG} to be the most influential one, producing the maximum of 27 % of ideal waveforms at the ψ_{CVG} values between 0° and 5°, regardless of the input values of other parameters (Figure 3e). An even more influential factor, however, was a combined parameter, i.e., the difference between the upper and lower margin oscillatory amplitudes A_{LUrel} defined as:

$$A_{LUrel} = (A_L-A_U)/(A_L+A_U)$$

When the lower and upper margin amplitudes were similar and the amplitude differences A_{LUrel} were around zero, the ideal waveforms occurred in 30 % of cases, regardless of the input values of other parameters (Figure 3f). The other individual input parameter values also showed an influence on the occurrences of ideal waveforms, but their maxima were smaller (20, 25, 24 and 20 % for A_U, A_L, H_W and φ_{VPD}, respectively, see Figure 3a-d). Nevertheless, since no individual parameter subranges exist in which ideal waveforms occur in 100% cases, favorable values of all the parameters should be combined to produce the ideal waveforms.

In 88 % of the cases, the waveforms did not fulfill the ideal conditions. In these cases, either the lower margin was hidden during some portion of the closing phase, or the upper margin was not at glottal edge during some portion of the opening phase. This makes the influencing factors on the waveform shape more complex and decreases interpretability of the VPD estimations for the non-ideal waveforms, e.g., when using sinusoidal fitting methods [4]. Therefore, we consider it useful to quantify the degree of "ideality" of the glottal waveforms when evaluating mucosal pliability and estimating the VPDs.

Acknowledgements:

We appreciate the help of R. Sandhanakrishnan in programming the code for generating and analyzing the waveforms. The work was supported by the Internal Grant Agency of the Palacký University project IGA_PrF_2022_029.

References:

[1] S. P. Kumar and J. G. Švec, "Kinematic model for simulating mucosal wave phenomena on vocal folds," *Biomed. Signal Process. Control,* vol. 49, pp. 328-337, 2019, doi: 10.1016/j.bspc.2018.12.002.

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

[3] S. Li, R. C. Scherer, M. X. Wan, and S. P. Wang, "The effect of entrance radii on intraglottal pressure distributions in the divergent glottis," *J. Acoust. Soc. Am,* vol. 131, no. 2, pp. 1371-1377, 2012, doi: 10.1121/1.3675948.

[4] J. J. Jiang, Y. Zhang, M. P. Kelly, E. T. Bieging, and M. R. Hoffman, "An automatic method to quantify mucosal waves via videokymography," *Laryngoscope,* vol. 118, no. 8, pp. 1504-1510, 2008, doi: 10.1097/MLG.0b013e318177096f.

HOW DO MULTIPLE SOUND SOURCES IN THE AIRWAY INTERACT?

Ingo R. Titze
Utah Center for Vocology, University of Utah, Salt Lake City, Utah, U.S.A.

Keywords: Multiphonics, Sound Sources, Voice Simulation, Sound Production

Abstract:

Objectives/Introduction:
The objective of this presentation is to demonstrate the nature of interaction between multiple self-oscillating sound sources distributed serially in an airway. Potential combinations are the true folds, the false folds, the aryepiglottic folds, the velum, the tongue, and the lips.

Methodology:
The methodology is computational. A simplified Navier Stokes solution is used to solve air transport and wave propagation from the base of the trachea to the lips. Various sections along the airway are selected as vibratory sound sources by changing the viscoelastic wall properties and producing a “fold-like” geometry with several sections.

Results:
Results indicate that serially coupled oscillators tend to entrain one another with aero-acoustic energy transfer. Their frequencies, amplitudes and phases are not independent. Glottal configuration, proximity of the sources to each other and to positively reactive airway segments, are major components of source dominance.

Conclusions:
It is concluded that these results may have bearing on multi-phonic sound production in singing, voiced consonant production in speech, vocalization with semi-occluded vocal tracts, and snoring.

Acknowledgements:

Research reported in this presentation was supported by National Institute on Deafness and Other Communication Disorders of the National Institutes of Health under grant number 1R01DC017998.

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References:
None.

SOURCE-FILTER COUPLING IN CONNECTED SPEECH SAMPLES

Lynn Maxfield[1], Brad Story[2], Anil Palaparthi[1,3], Sarah Hargus Ferguson[4], Ingo R. Titze[1,3]

[1] Utah Center for Vocology, University of Utah, Salt Lake City, UT, USA
[2] Department of Speech, Language, and Hearing Science, University of Arizona, Tucson, AZ, USA
[3] National Center for Voice and Speech, Salt Lake City, UT, USA
[4]Department of Communication Sciences and Disorders, University of Utah, Salt Lake City, UT, USA

***Keywords*: Acoustics, Source-Filter Interaction, Formant-Harmonic Synchronization**

Abstract:

Objectives / Introduction:

When humans produce speech for communication, often two contradictory objectives are targeted at once. If the first objective is to optimize intelligibility, then a low, relatively stable, fundamental frequency is desirable. This condition, with densely spaced harmonics and little entrainment between the sound source and the acoustic filter of the vocal tract allows for rapid and precise modulation of the sound source by the speech articulators. If the objective, however, is to optimize total radiated sound intensity and tone quality, then a variable fundamental frequency that allows for frequent and prolonged alignment of harmonics with resonances of the vocal tracts would be preferrable.

The purpose of this study is to investigate the degree to which these couplings between the oscillating sound source and the vocal tract filter occur in connected speech samples, helping to illustrate how humans may choose to optimize their vocalizations for intelligibility, intensity, or both.

Methods:

Data Collection - This study utilized both simulated voice signals and connected speech samples from human subjects. Simulated signals were produced using *TubeTalker* under two conditions. First, the fundamental frequency (f_o) was fixed at 100 Hz, while the cross-sectional areas of the articulatory space of the simulator were randomly varied to produce an approximation of all possible vowel conditions, in essence disallowing source-filter coupling. In the second and third conditions, the f_o was allowed to vary along with the articulatory space with the condition that $2f_o$ or $3f_o$ had to match the first resonance frequency of the resulting vocal tract shape, essentially forcing a complete coupling of the source and filter.

Speech samples were collected from 41 adult American English speakers (21 Female, 20 Male) as part of a separate study (Ferguson, 2004). Subjects were each recorded speaking 188 sentences. Sentences were categorized into 3 types: 1) /b *vowel* d/ sentences, with ten vowels in the /bvd/ context, each set in seven sentence frames – 70 sentences total, 2) consonant-vowel-consonant sentences, with words chosen or adapted from the Northwestern University Auditory Test, No.6 (NU-6; (Tillman & Carhart, 1966)), each set in two sentence frames – 104 sentences total, and 3) 14 additional sentences from the Central Institute for the Deaf (CID) Every Day Sentences test (Davis & Silverman, 1978). The sentences from each subject were trimmed and concatenated, resulting in 41 samples, each containing between seven and ten minutes of connected speech.

Analysis – for all samples, simulated and human, acoustic signals were analyzed to compute a set of three "density" plots. A Voice Range Density normalized (VRD_N) indicates the density of the *SPL* and f_o values, with color intensity indicating the amount of time spent in that SPL/f_o combination. A Vowel Space Density normalized (VSD_N) shows the density of the F_1 and F_2 formant frequencies. Finally, the Source-Tract Coupling Density (STCD) indicates the population of f_o/F_1 pairs, with the white lines indicating where $F_1=nf_o$, indicating how much time is spent during the sample with alignment between various harmonics (nf_o) and F_1.

Results:

Density plots for the simulated samples provide compelling visualization of the extreme conditions, where source filter couple is either non-existent, or nearly constant. In human subjects, there is naturally more variability between subjects. Still, in large speech samples, the density plots indicate that most speakers spent considerable amounts of their phonation time with a coupling between the source and filter as visualized by the STCD.

Conclusions:

This study provides evidence that, in the absence of specific instructions regarding voice production, humans can optimize their vocalization to align/couple formant frequencies with harmonics of the sounds source. Given the prevalence of these couplings across subjects and the lack of instruction regarding voice quality targets, it seems unlikely that the subjects were consciously deciding to make these adjustments. Instead, it appears plausible that the

voice system self-organized to optimize beneficial interactions between the source and filter. Future research will investigate if and how subjects may optimize their vocalizations in response to external directions to target their voice production (e.g. speak loudly, or speak clearly).

Acknowledgements:

Research reported in this presentation was supported by National Institute on Deafness and Other Communication Disorders of the National Institutes of Health under grant number 1R01DC017998.

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References:

Davis, H., & Silverman, S. R. (1978). *Hearing and Deafness, 4th Ed.* New York: Holt, Rinehart, and Winston.

Ferguson, S. H. (2004). Talker differences in clear and conversational speech: Vowel intelligibility for normal-hearing listeners. *The Journal of the Acoustical Society of America, 116*, 2365-2373.

Tillman, T., & Carhart, R. C. (1966). *An expanded test for speech discrimination utilizing CNC monosyllabic words: N.U. Auditorisy Test No. 6.* USAF School of Aerospace Medicine Report No. SAM-TR-66-55.

Influence of subglottal vibrational excitation in the range of 2-6 kHz on voice parameters

Patrick Hoyer[1], Pedro Amarante Andrade[2], Marek Frič[3]

[1] Fraunhofer Head Department, Munich, Germany

[2] Curtin School of Allied Health, Faculty of Health Sciences, Curtin University, Australia

[3] Musical Acoustics Research Centre, Academy of Performing Arts in Prague, Czechia

***Keywords*: subglottal resonances; vocal training; vibrational stimulation**

Abstract:

Objectives / Introduction:

Vocal activity is primarily perceived as an acoustic phenomenon. However, vibrational sensations at different parts of the body of the speaker/singer also emerge. The airways in the vocal tract and the subglottal system are both vibrationally excited during phonation and these vibrational sensations are essential for trained singers to perform in harsh acoustical environments on stage. The impact of subglottal resonances has been discussed in literature (Hanna et al. 2018; Tietze 2023; Hoyer et al. 2022). The vibrational amplitude measured at the jugulum (soft region between the cricoid cartilage and the sternum) is an indicator for vocal settings (Frič et al. 2024) and correlates with basic vocal parameters as the sound pressure level (SPL) as well (Sveč et al. 2005). In contrast to supraglottal resonances, subglottal resonances are less flexible and little interaction has been found for the first two vocal tract resonances. However, at frequencies above 2 kHz, vocal tract resonances may cross the vocal folds that serve as an efficient barrier between upstream and downstream acoustics at lower frequencies. Simulations based on the resonance structure of the total airway suggest the importance of the upper region of the windpipe just below the glottis for the vocal output (Titze 2023).

While an external vibrational excitation has been used to estimate vocal tract resonances (Pham Thi Ngoc and Badin 1994), to our knowledge no attempt has been made to measure the vocal parameters after vibrational stimulation of the subglottal region. As mechanical interactions are strong at frequencies below 800 Hz, we chose frequencies between 2 and 6 kHz which are relevant for the sound quality but where vibratory interactions are limited (Sundberg 1992).

Methods:

The vibrational excitation was induced at the skin at the sternum using a Dynavox EXC-25 (Sintron Distribution GmbH, Sinsheim, Germany) exciter (excitation voltage 2.9 Vpp). The frequency range of the broadband vibrational signal ranged from 2 kHz to 6 kHz. A headband fixed microphone was placed at 30 cm distance (Sennheiser MKE2, Germany) and an EGG (Laryngograph D-200, icSpeech, Kent UK) was used with an accelerometer (PCB Electronics 352C23) placed at the EGG electrode surface.

After warm-up, a scale singing in medium dynamic in the whole pitch range on vowel [a] as well as all prolonged vowels in 3 pitches (D4, A4, D5), a simple song (happy birthday) and a standard reading performed by four female singers before and after the intervention were recorded. The singers were asked to phonate with the exciter pressed to their chest. During the training with excitation, the participants were asked to phonate vowels [i:, e:, a:, o:, u:] in common Check pronunciation and in a comfortable range at different pitches as well as pitch glides for a duration of up to 5 min.

The vocal data were extracted using the Praat software (Boersma and Weenink, Amsterdam, Netherlands; Version 6.3.17). The EGG signal was evaluated using the DECOM method implemented in MATLAB (Henrich et al 2004).

Results:

The treatment with a broadband vibrational excitation between 2 and 6 kHz at the chest (sternum) induced changes in the voice of the participants after the intervention. The reported changes after the intervention of the four participants were coherent in the categories “more resonant” and “increase in volume” in respect to their voice.

Acoustical findings confirm an increasing SPL and a rise in the first harmonic as well as a rise in higher harmonics (CPPS). The decrease of noise-to-harmonic ratio and the frequency position of the third formant in [i:, e:, a:, o:, u:] vowels (paired t test) were observed.

In scale singing ([a:] vowel) the results showed similarly an increase of SPL and the first harmonic level increased as well. As for CPPS, the spectral tilt, spectral slope, and spectral centroid (0 - 2.5 kHz) increased. Furthermore, the frequency position of the first, the second, and the fifth formant were raised. The contact coefficient decreased when compared with respect to f0 and SPL for both speaking voice and singing scales, whereas the singing power ratio (SPR) generally increased for singing scales.

Conclusions:

A vibrational excitation at the sternum in the frequency range of 2 kHz to 6 kHz has an impact on the voice of the four female participants. A lower contact coefficient together with an increase in SPL indicates that the voice efficiency is increased after the intervention. The excitation of the subglottal airways during singing may influence the connectivity between the subglottal and the supraglottal region. The influence of the intervention may be due to a training effect, an alternation of the glottal flow and/or a stimulus of the tissue. Further research is necessary to elucidate the underlying process of the intervention further.

Acknowledgements:

The authors want to thank the participants of the study. This publication was supported by the Institutional Endowment for the Long-Term Conceptual Development of Research Institutes at the Academy of Performing Arts in Prague.

References:

Frič M., Dobrovolná A., Andrade P. A. Comparison of laryngoscopic, glottal and vibratory parameters among Estill qualities – Case study. In: *Biomedical Signal Processing and Control* 87(2024):1-19 DOI: 10.1016/j.bspc.2023.105366

Hanna, N.; Smith, J.; Wolfe, J.: How the acoustic resonances of the subglottal tract affect the impedance spectrum measured through the lips. In: *J Acoust Soc Am* 143 (5) (2018): 2639–2650 DOI: 10.1121/1.5033330.

Henrich N, d'Alessandro C, Doval B, Castellengo M. On the use of the derivative of electroglottographic signals for characterization of nonpathological phonation. In: *J Acoust Soc Am.* 115 (3) 2004 :1321-1332. doi:10.1121/1.1646401

Hoyer, P.; Riedler, M.; Unterhofer, C.; Graf, S.: Vocal Tract and Subglottal Impedance in High Performance Singing: A Case Study. In*: Journal of Voice* (in press). DOI: 10.1016/j.jvoice.2022.01.015.

Pham Thi Ngoc, Y., and Badin, P. Vocal tract acoustic transfer function measurements: further developments and applications. J. Phys. *IV France 04* (1994), C5-549-C5-552. DOI: 10.1051/jp4:19945118.

Sundberg, Johan: Phonatory Vibrations in Singers: A Critical Review. In: *Music Perception* 9 (3) 1992, 361–381. DOI: 10.2307/40285557.

Sveč, Jan G.; Titze, Ingo R.; Popolo, Peter S.: Estimation of sound pressure levels of voiced speech from skin vibration of the neck. In: *J Acoust Soc Am* 117 (3), (2005): 1386–1394. DOI: 10.1121/1.1850074.

Titze, Ingo R.: Simulation of Vocal Loudness Regulation with Lung Pressure, Vocal Fold Adduction, and SourceAirway Interaction. In: *Journal of voice* 37 (2) (2023): 152–161. DOI: 10.1016/j.jvoice.2020.11.030.

INTRALARYNGEAL NEURAL NETWORKS, A NOVEL AVENUE IN NEUROLARYNGOLOGY AND ITS IMPACT ON PHONATORY POSTURING

Tracicaru RV[1,2], Hînganu MV[1], Hînganu D[1], Bräuer L[2], Friedrich P.[2]

[1] Department of Morphofunctional Sciences 1, Chair of Anatomy and Embryology, "Gr. T. Popa" University of Medicine and Pharmacy, Iasi, Romania

[2] Institute for Functional and Clinical Anatomy, Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany

***Keywords*: Neuromuscular system, Larynx, Reflex, Electron microscopy**

Abstract:

Objectives / Introduction:

In the current understanding of speech production, it is widely believed that the cricothyroid and thyroarytenoid muscles are mainly responsible for fundamental frequency control. (1) However, previous studies have shown that the correlation between muscular activity and fundamental frequency is not perfect and, furthermore, is not always predictable. (2) The particularities of the laryngeal neuromuscular system have long been reported (3–7), and other dedicated muscle groups also show similar particularities (e.g. the extraocular eye muscles), but no consensus has yet been reached. Our study aims to provide new insights into the neural control of laryngeal muscles that may explain some of the functional peculiarities of the intrinsic laryngeal muscles during posture and vocalization.

Methods:

Two complementary studies were conducted on porcine and human larynges to determine the functional and anatomical characteristics of the intrinsic laryngeal muscles and their innervation. Larynxes from 20 pigs were freshly excised and investigated electromyographically. For this purpose, the cricothyroid or thyroarytenoid muscle was stimulated with direct current (in the absence of external innervation) and a possible reaction in the respective opposite muscle pair was recorded. To confirm the functional results, 10 human larynges were dissected and processed for two-color immunofluorescence with neurofilament 70 KD and either synaptophysin or ACtH and for transmission electron microscopy. For immunofluorescence, two sections were selected from each hemi-larynx, while the corresponding hemi-larynx was processed for electron microscopy. For this purpose, two small muscular biopsies were taken from the vocalis muscle in its anterior and posterior half near the vocal ligament. Ultra sections were selected from each biopsy, usually 2 per biopsy fragment.

Results:

Electromyography revealed simultaneous contraction of the cricothyroid and thyroarytenoid muscles in all cases. The contraction was bidirectional and had similar amplitude in both muscles. This indicates a reflex pathway between the two muscles that is not related to the classical innervation of the larynx, as the classical innervation was severed during laryngeal excision. At the same time, both the immunofluorescence and electron microscopy revealed the presence of neurons distributed between the muscle fibers within the vocalis muscle, suggesting the presence of a neural network and a possible center for the reflex pathway described previously. Furthermore, the muscle fibers of the larynx appear to be multiply innervated, as has been formulated by other authors, and their muscle structure differs from that of other striated muscular fibers in the human body. The presence of an agonist-antagonist reflex and the discovery of nerve bodies within the vocalis muscle suggests that the larynx has a local neuronal network that is independent of external innervation by the superior and inferior laryngeal nerves. This in turn influences how the interactions between the cricothyroid and thyroarytenoid muscles are modeled especially with regard to the tension of the body and cover layers of the vocal fold during phonation.

Conclusions:

In addition to innervation via the superior and inferior laryngeal nerves, the larynx has a neuromuscular system with complex intrinsic innervation and specific muscle fibers. Our study emphasizes both the morphological concept for the network and its electromyographic features. The role of the network needs to be further analyzed. The network allows the cricothyroid and cricoarytenoid muscles to control each other's activation. This explains the imperfect correlation and allows for greater complexity in speech gestures. Further development of this study is needed with an increase in the sample size and a systematic analysis of the vocal folds to fully characterize the neural network. To better understand

the discovered reflex, studies on human subjects with previous laryngeal nerve lesions could highlight its functioning during vocal gestures. These initial results could have a lasting impact on our understanding of muscular activation patterns during vocalization.

Acknowledgements:

There are no conflicts of interest to report. This study was conducted as part of the first author's joint PhD program (doctorate), which is a collaboration between the "Gr. T. Popa" University of Medicine and Pharmacy, Iasi, Romania and the Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany.

References:

1. Titze IR. Principles of Voice Production. National Center for Voice and Speech; 2000. 448 p.

2. Poletto CJ, Verdun LP, Strominger R, Ludlow CL. Correspondence between laryngeal vocal fold movement and muscle activity during speech and nonspeech gestures. J Appl Physiol Bethesda Md 1985. 2004 Sep;97(3):858–66.

3. Goerttler K. Die Anordnung, Histologie und Histogenese der quergestreiften Muskulatur im menschlichen Stimmband. Z Für Anat Entwicklungsgeschichte. 1951 Jul 1;115(4):352–401.

4. Neuhuber WL, Wörl J, Berthoud HR, Conte B. NADPH-diaphorase-positive nerve fibers associated with motor endplates in the rat esophagus: new evidence for co-innervation of striated muscle by enteric neurons. Cell Tissue Res. 1994 Apr;276(1):23–30.

5. Chhetri DK, Neubauer J, Berry DA. Neuromuscular control of fundamental frequency and glottal posture at phonation onset. J Acoust Soc Am. 2012 Feb;131(2):1401–12.

6. D'Antona G, Megighian A, Bortolotto S, Pellegrino MA, Marchese-Ragona R, Staffieri A, et al. Contractile properties and myosin heavy chain isoform composition in single fibre of human laryngeal muscles. J Muscle Res Cell Motil. 2002;23(3):187–95.

7. Hisa Y. Neuroanatomy and Neurophysiology of the Larynx. Springer Japan; 2016. Available from: https://books.google.de/books?id=fNh2DQAAQBAJ

EXPLORING NONLINEAR PHENOMENA (NLP) IN ANIMAL VOCALISATIONS THROUGH OSCILLATOR THEORY

Hanspeter Herzel[1]

[1] Institute for Theoretical Biology, Charité and Humboldt University, Berlin, Germany

Abstract:

It was discovered more than 20 years ago that animal vocalizations often exhibit nonlinear phenomena, such as subharmonics, biphonation, and deterministic chaos. Since then, researchers have collected and classified hundreds of calls from various animals, including monkeys, deer, dogs, horses, farm animals, and zebra finches. These recordings provide valuable information about sound production mechanisms and the potential functions of nonlinear phenomena.

Advanced oscillator theory can help analyze these complex data sets. Bifurcations and attractors can be visualized using "Arnold tongue diagrams," representing bifurcation diagrams with frequency ratios on the x-axis and coupling strength on the y-axis. The asymptotic dynamics ("attractors") can be classified as frequency locking, torus, or deterministic chaos for any set of parameters. Slowly varying parameters, such as frequencies, sound intensity, or coupling strength, induce well-defined qualitative changes known as "bifurcations."

The combination of data analysis, mathematical modeling, and oscillator theory can lead to testable predictions. For instance, additional vibratory structures, such as vocal membranes, can increase the occurrence of nonlinear phenomena. Varying parameters should result in a sequential ordering of Arnold tongues and tori called the "devil's staircase." The prevalence of biphonation in horse vocalizations suggests weak coupling of two sound-generating mechanisms. The synergy between data analysis and oscillator theory can contribute to a better understanding of the biological functions of nonlinear phenomena.

DOES SEX MATTER? IN VITRO EXPOSURE OF VOCAL FOLD FIBROBLASTS TO VIBRATION

Andrijana Kirsch[1,2], Tanja Grossmann[1], Barbara Steffan[1], Andrea Groselj-Strele[3], Claus Gerstenberger[1], Markus Gugatschka[1]

[1] Division of Phoniatrics, Department of Otorhinolaryngology, Medical University of Graz, Austria
[2] current affiliation: Division of Physiology and Pathophysiology, Otto Loewi Research Center for Vascular Biology, Immunology and Inflammation, Medical University of Graz, Austria
[3] Core Facility Computational Bioanalytics, Center for Medical Research, Medical University of Graz, Austria

***Keywords*: vocal fold fibroblasts; vibration; sex-specific**

Abstract:

Objectives / Introduction:
Homeostasis in healthy human vocal folds is maintained by complex interactions between the protective epithelium and the underlying fibroblasts within the lamina propria. Vocal fold fibroblasts synthesize most of the extracellular matrix components in response to signals from the surrounding environment. The differentiation of the native lamina propria into a unique multilayered tissue during human development is correlated with the physiological vibration of the tissue. This research focused on molecular changes in vocal fold fibroblasts driven by sex-specific vibration patterns using a phonomimetic vibration device established by our group (Kirsch et al. 2019).

Methods:
A potential sex-specific voice frequency effect was investigated in an in vitro setting using immortalized human vocal fold fibroblasts from a male and a female donor. Cells and supernatants were harvested after exposure to either male or female vibration frequency with or without the addition of cigarette smoke extract. Gene and protein analysis was performed by means of qPCR, Western blot, ELISA and Luminex.

Results:
We found that exposure of cells to both male and female vibration patterns did not induce significant changes in the expression of extracellular matrix-, inflammation-, and fibrosis-related genes compared to control cells. The addition of cigarette smoke extract to vibration downregulated the gene expression of *COL1A1* in cells exposed to the female vibration pattern, and induced *MMP1* and *PTGS2* in cells exposed to both female and male vibration patterns. The protein expression of MMP1 and COX2 was found to be significantly regulated only in cells exposed to cigarette smoke extract and female vibration pattern.

Conclusions:
Different vibration patterns alone did not cause different cell responses in our experimental setting. However, when cigarette smoke extract was added, the female vibration pattern had a tendency to elicit/maintain more pro-inflammatory responses in exposed cells than the male vibration pattern.

Acknowledgements:
Thanks to Jennifer Ober for her excellent technical support.

References:
Kirsch A, Hortobagyi D, Stachl T, Karbiener M, Grossmann T, Gerstenberger C,Gugatschka M (2019) Development and validation of a novel phonomimetic bioreactor. PLoS One 14(3):e0213788.

METABOLIC VALIDATION OF AN ANAEROBIC VOCAL DEMAND TASK: NOVEL APPLICATION OF BLOOD LACTATE AS A BIOMARKER OF VOCAL FUNCTION

Mariah E. Morton-Jones[1], Andreas N. Kavazis[1], Mary J. Sandage[2]

[1] School of Kinesiology, Auburn University, Auburn University, Auburn, AL, USA
[2] Department of Speech, Language, and Hearing Sciences, Auburn University, Auburn, AL, USA

***Keywords*: Voice Physiology; Metabolic Biomarker; Blood Lactate; Metabolism**

Abstract:

Objectives / Introduction: Voice production entrains intrinsic laryngeal skeletal muscles as well as skeletal muscles of the respiratory and articulatory systems, the metabolism of which is not well understood in response to performance demands. To date, voice science has yet to establish quantifiable metabolic biomarkers of vocal aptitude elucidated from a vocal demand task to better characterize the training response of the speech musculature (respiratory, laryngeal, articulatory) to behavioral voice training programs that are designed to optimize vocal function and mitigate vocal fatigue, a commonly reported symptom among occupational voice users. Given current primary reliance on indirect measures of vocal function (acoustic, aerodynamic, perceptual) in research and clinic, development of an objective biomarker of human laryngeal tissue response *in vivo* is compelling.

Muscle biopsy routinely used in exercise physiology to quantify and characterize muscle metabolism is not viable for study of voice physiology, given the small size of the intrinsic laryngeal skeletal muscles compared to limb muscle as well as their inaccessibility and critical role in biological function (airway protection, breathing, swallowing, and speech). An alternative approach that is novel and innovative to study vocal function is quantification of metabolic substrates in blood samples pre and post vocal demand tasks. Blood lactate concentration is commonly used to assess metabolic demand and skeletal muscle training response. La^- is a by-product of the glycolytic energy system of anaerobic metabolism, which synthesizes adenosine triphosphate (ATP) and fuels performance of skeletal muscle contraction and relaxation for activity, including speech production. This investigation sought novel application of exercise science methodology to determine if a change in blood lactate was detectable in an anaerobically designed vocal demand task ($VCAT^{60}$) and establish construct validity of the developed vocal demand task in assessing the anaerobic capacity of the speech musculature, similar to anaerobic power tests commonly used in applied exercise science.

Methods: A prospective, multimodality repeated measures study quantified blood lactate concentration pre and post-$VCAT^{60}$ in 51 vocally healthy adults. The secondary outcomes included determining correlations and predictors of the change in lactate including: aerodynamic and anthropometric data as well as participant reported vocal fatigue.

Results: A significant change in lactate pre and post $VCAT^{60}$was observed ($p = 0.003$).

Conclusions: Evidence supports the hypothesis that blood lactate may be an effective metabolic biomarker in investigating human voice physiology *in vivo*. The $VCAT^{60}$ provides a means from which to understand metabolic adaptations that occur in the peripheral muscles of the speech mechanism, a valuable metric to further study vocal effort and vocal fatigue. This is the first study of its kind to establish construct validity of an anaerobic vocal demand task within a metabolic framework *in vivo* via blood lactate, which may have several impactful implications in voice science and clinical care.

References:

Ekblom, B. (2017). The muscle biopsy technique. Historical and methodological considerations. *Scandinavian journal of medicine & science in sports, 27*(5), 458-461.

Ferguson, B. S., Rogatzki, M. J., Goodwin, M. L., Kane, D. A., Rightmire, Z., & Gladden, L. B. (2018). Lactate metabolism: historical context, prior misinterpretations, and current understanding. *European journal of applied physiology, 118*(4), 691-728.

Gladden, L. B. (2010). Lactate transport and exchange during exercise. *Comprehensive Physiology*, 614-648.

Morton-Jones, M. E., Gladden, L. B., Kavazis, A. N., & Sandage, M. J. (2023). A Tutorial on Skeletal Muscle Metabolism and the Role of Blood Lactate: Implications for Speech Production. *Journal of Speech, Language, and Hearing Research*, 1-15.

Snell, E. N., Plexico, L. W., Weaver, A. J., & Sandage, M. J. (2020). Quantifying vocal power: correlation of whole-body anaerobic power to vocal function measures. *Journal of Speech, Language, and Hearing Research, 63*(8), 2597-2608.

Sandage, M. J., & Pascoe, D. D. (2010). Translating exercise science into voice care. *Perspectives on voice and voice disorders, 20*(3), 84-89.

SINGLE-CELL ATLAS OF VOCAL FOLD AND LARYNGEAL EMBRYOGENESIS, MATURATION AND AGING

Tadeas Lunga[1], Elliot Xie[2], Vlasta Lungova[1], Chitrasen Mohanty[2], Christina Kendziorski[2], Susan Thibeault[1]

[1] Division of Otolaryngology – Head and Neck Surgery, University of Wisconsin - Madison, Madison, WI, USA
[2] Department of Biostatistics & Medical Informatics, University of Wisconsin - Madison, Madison, WI, USA

***Keywords*: Voice; Tissue Engineering; Transcriptomics; Development, Larynx**

Abstract:

Objectives / Introduction:

Embryogenesis, maturation, and aging of the larynx and vocal folds (VF) are complex processes. In the early stages, during embryonic and fetal development, the larynx and VF undergo shaping and the initiation of functional roles. Postnatally, growth and maturation persist in the laryngeal and VF structures. Structural differences of VF mucosa have been observed across various age groups, encompassing newborns, adolescents, adults, and elderly. Particularly, the elderly exhibit degenerative changes in VF mucosa.

These structural differences in VF mucosa across lifespan predispose individuals to age-associated voice disorders. For instance, newborns often manifest voice disorders linked to impaired embryogenesis/fetal development, while the elderly may present with presbyphonia attributed to VF degenerative changes. Despite the clinical relevance of these age-associated manifestations, treatment options are limited, primarily due to a deficient understanding of the foundational cellular and genetic processes driving these structural changes.

To address our critical knowledge gaps, our study aimed to identify and comprehend the cellular heterogeneity of VF across the lifespan. Through the application of the cutting-edge technique of single-cell RNA sequencing, we elucidated the key cellular components and molecular pathways within VFs and created the first single-cell atlas of VF and laryngeal embryogenesis, maturation, and aging. This approach serves as a pivotal step toward advancing therapeutics tailored to the specific needs of different age groups, ultimately enhancing our understanding and clinical management of agerelated voice disorders.

Methods:

The Bl6 mouse strain was used for this investigation. Murine larynges were harvested at embryonic (E) stages E11.5, E13.5, E15.5, and E18.5; eight larynges were pooled together for one biological replicate, with three biological replicates collected for each time point. In postnatal (P) stages, larynges were collected at birth (P0), 4, 12 weeks, 1 and 1.5 years. For P0, six larynges were pooled for each biological replicate; for other postnatal stages, six VF tissue pairs were pooled. Three biological replicates were used for each time point. Harvested tissues were enzymatically processed into singlecell suspensions and used for cDNA library preparation subsequently sequenced on NovaSeq6000. The sequenced samples underwent quality control and read alignment. Data preprocessing steps normalization, integration, and clustering together with annotations were done in Seurat. Erythrocytic and thymotic clusters were removed from the analysis. Annotations of individual cell clusters were based on the top 20 differentially expressed genes (DEG) and using Cellkb. Enrichment analysis of individual cell clusters for top biologic processes, ontology, and transcription factors was performed with Enrichr - Ma'ayan Laboratory, Computational Systems Biology.

Results:

Unsupervised clustering across all stages resulted in a UMAP identifying 48 clusters (**Figure 1**). All cell clusters were established early in development. These clusters comprised major cell populations such as epithelial, mesenchymal, endothelial, immune, muscular, and neuronal (**Figure 1**). Within the epithelial cluster, basal cell progenitors, basement membrane-producing cells, two subtypes of mature basal cells, three types of suprabasal cells, two types of ciliated epithelial, and four types of secretory epithelial cells were identified. In the mesenchymal cluster, three types of proliferating mesenchymal progenitors, seven subtypes of fibroblasts originating from lateral and neural crest cell sources, chondroblasts, chondrocytes, two types of smooth muscle cell clusters, myoblasts, and three types of skeletal muscle cell clusters were distinguished. Among the immune cell clusters, five types of macrophages, dendritic cells, neutrophils, and lymphocytes were observed. Additionally, four types of endothelial cells were identified. Lastly, two clusters of neural crest cell-derived (NCC) cell types such as neurons and Schwann cells were revealed.

The complexity of cell populations underwent a progressive increase from embryonic stages to adulthood (P12wks). Cell complexity reached its peak in between adolescence and adulthood (4 to 12wks) and subsequently declined. In particular, embryonic stages (E11.5 – 15.5) were enriched mostly for mesenchymal clusters - proliferating mesenchymal

progenitors and chondroblasts. The highest cell count for basal cell progenitors was measured at E11.5. Endothelial clusters appear at E15.5. Postnatal stages from P0 to P12wks predominantly displayed an increase in epithelial clusters. Epithelial cluster complexity decreased, specifically the reduction of secretory cells from P1year of age to P1.5years of age. Immune cells, predominantly macrophages and neutrophils, had a steady rise at late E18.5 peaking at P1.5.

The major molecular pathways activated within fibroblasts were extracellular matrix (ECM) organization, collagen fibril organization, and extracellular structure organization. The major transcription factors present in this subtype were SMAD3, GFI1B, SMAD1, SMAD4 and STAT4.

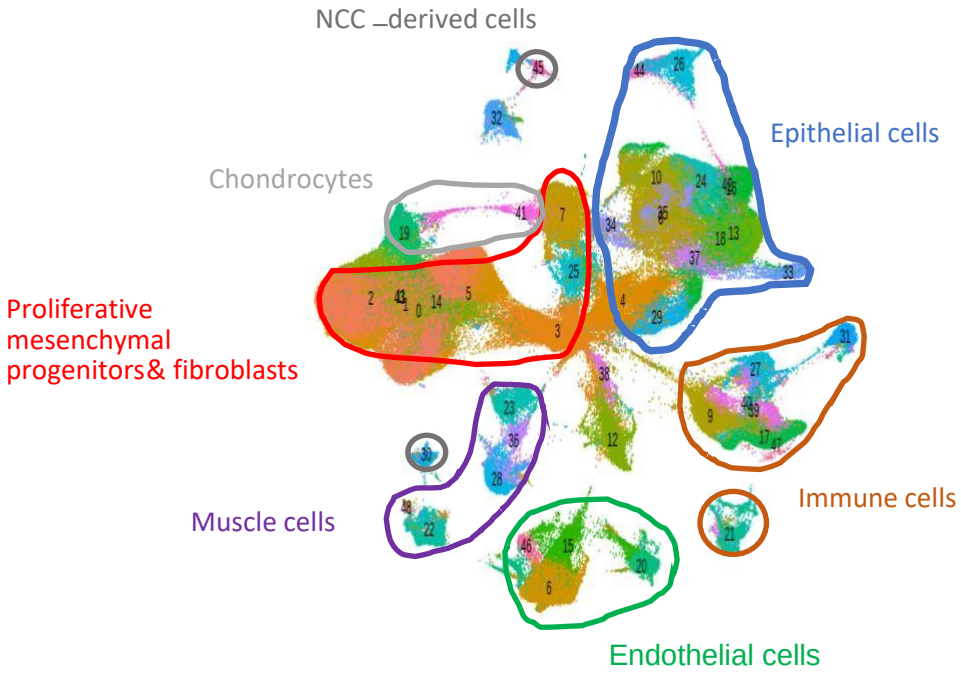
Conclusions:

The single-cell atlas presented in this study sheds light on the cellular heterogeneity of VF and laryngeal tissue throughout development. Major VF cell populations are established early in development. Cell complexity increases until adulthood with **cell proliferation** and **maturation,** however, decreases in later stages of development (1 - 1.5 years of age) with laryngeal/VF tissue **degeneration.** Early embryonic stages are accompanied by differentiation of laryngeal cartilages and VF lamina propria supported by predominantly mesenchymal clusters and activation of genes responsible for ECM and collagen fibril organization. Maturation of VF epithelium is apparent from P0 to 12wks of age supported by an increase in epithelial clusters. Late postnatal stages are characterized by tissue degeneration specifically reduction in epithelial cell complexity (secretory cells) and with an influx of immune cells, specifically macrophages and neutrophils.

Acknowledgements:

The author(s) utilized the University of Wisconsin – Madison Biotechnology Gene Expression Center (Research Resource Identifier - RRID:SCR_017757) for RNA library preparation and the DNA Sequencing Facility (RRID:SCR_017759) for sequencing. This work is funded by NIDCD R01 DC0433618 and GM102756.

Figure 1:



MINIMALLY INVASIVE IN SITU BIOPRINTING FOR VOCAL FOLD REGENERATION USING A CABLE-DRIVEN CONTINUUM MANIPULATOR

Swen Groen[1], Audrey Sedal[1], Luc Mongeau[1]

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

***Keywords*: In situ bioprinting; Continuum robotics; Vocal fold surgery, New technologies**

Abstract:

Objectives / Introduction:

Hydrogels are currently used for soft tissue regeneration and repair[1, 2]. Hydrogel injections are possible for a class of shear thinning materials. Local injections tend to be challenging due to lack of spatial control and can complicate procedures in cases of deteriorated epithelium[3]. *In situ* bioprinting is a novel technology to decrease complications for injections and soft implantable materials. *In situ* bioprinting is the deposition of bio-inks, amongst which hydrogels create three-dimensional constructs directly inside the human body instead of through the creation of grafts or *ex vivo* printed constructs[3, 4]. *In situ* bioprinting may offer improved accuracy for the deposition of hydrogel in open wounds[5]. In applications such as skin and bone, current in situ bioprinting methods require open surgery. Currently, no methods exist to perform *in situ* bioprinting in vocal fold tissue. Here, we investigated a novel design for in situ bioprinting, using a minimally invasive method. The device was designed for use in vocal fold surgical procedures through an endoscope. The ultimate goal is to reduce complications during vocal fold injections. The device uses a cable-driven continuum manipulator to deposit hydrogels. During deposition control of the manipulator is possible, in order to create desired geometries. To create accurate control, a kinematic model has been developed based on the geometric deformation due to actuation. A handheld device enables position control of the deposition. A prototype was built for in situ bioprinting onto the vocal folds. The kinematic model was verified for accurate positioning and printability of Pluronic F217 was investigated. The results show a proof of concept for the use of continuum manipulators for *in situ* bioprinting.

Methods:

The proposed design of the manipulator consists of a hollow silicon cylinder, with internal wires for actuation. To decrease radial expansion during deposition, SLA printed disks are incorporated.

The continuum robot was manufactured using a SLA printer and a casting of silicone in printed molds. The derived kinematic model has been verified using object tracking. The manipulator is actuated along a patterned grid. Images are obtained at each point in the grid from a frontal and side view. To decrease the error the manipulator was pretensioned for specified values. A Hough filter was used to determine the tip location at each grid point and compare the calculated position. The optimal value of pretension for minimal error was determined for further tests. Hysteresis was tested by actuation of the manipulator for a specific location and the movement of the tip was measured at specified time intervals. Repeatability of control of the manipulator was measured by repeating a prescribed squared pattern for a total of 10 times. The tip location was determined at each point in time. The results were compared to the predicted motion for each individual cycle.

Printability of the manipulator was tested using a 15% w/v Pluronic F217 hydrogel. Printing was performed onto a heated substrate at a temperature of 37 degrees, resulting in crosslinking of Pluronic F217. During printing a fixed writing speed of 3 mm/s was used. Optimal extrusion speeds were obtained for two different nozzle diameters, 18. The optimal extrusion speed was determined by quantifying the average filament thickness and visual inspection of filament quality. Printability of objects was tested with both manual control and pre-determined printing paths.

Results:

The described manufacturing procedure was shown to have an average production error 0.2mm. The diameter of the manipulator is 8mm, which is smaller than any described *in situ* bioprinting methods so far. Verification of the kinematic model showed that an optimal pretension of 10% resulted in an average position error of 0.354±0.018 mm. The average error was determined over a work region of 1 cm radius, which is appropriate for vocal fold surgery. There was no significant impact of hysteresis after actuation. Repeatability of a prescribed printing path showed an average error of 0.13±0.05 mm between cycles.

The extrusion speeds for the used Pluronic F217 were optimized for optimal resolution, itself dependent on the diameter of the nozzle used. The optimal extrusion speed for the 18 gauge and 27 gauge nozzle were 59.5 and 51 µl/s respectively with a resolution of 2.3 and 0.92 mm. Lastly printability of constructs was confirmed with both predetermined printing paths and manual control during printing.

Conclusions:

Herein we describe a novel design for a cable-driven continuum manipulator, which is capable of extrusion of hydrogels. During deposition the position of the manipulator can accurately be controlled to create geometric constructs which mimic the vocal fold tissue. The printability of Pluronic F217 was demonstrated and optimized for resolution. Concept was illustrated by using pre-determined printing paths as well as manual printing.

Acknowledgements:

This work was supported by the National Institute of Health (RO1-DC018577). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References:

1. Wan-Chiew N, Baki MM, Fauzi MB, Lokanathan Y, Azman M. In Vitro Evaluation of Biomaterials for Vocal Fold Injection: A Systematic Review. Polymers. 2021;13(16).

2. Gaston J, Thibeault SL. Hyaluronic acid hydrogels for vocal fold wound healing. Biomatter. 2013;3(1):e23799.

3. Samandari M, Mostafavi A, Quint J, Memić A, Tamayol A. In situ bioprinting: intraoperative implementation of regenerative medicine. Trends in Biotechnology. 2022.

4. Singh S, Choudhury D, Yu F, Mironov V, Naing MW. In situ bioprinting – Bioprinting from benchside to bedside? Acta Biomaterialia. 2020;101:14-25.

5. Di Bella C, Duchi S, O'Connell CD, Blanchard R, Augustine C, Yue Z, et al. In situ handheld three-dimensional bioprinting for cartilage regeneration. J Tissue Eng Regen Med. 2018;12(3):611-21.

THE POTENTIAL ROLE OF SP-G AND PLUNC IN WOUND HEALING AND TUMOR PATHOGENESIS IN THE HUMAN LARYNX RESPECTIVELY THE VOCAL FOLD

Aurelius Scheer[1], Martin Schicht[1], Friedrich Paulsen[1], Markus Eckstein[2] Antoniu-Oreste Gostian[3], Lars Bräuer[1]

[1]Institute of Functional and Clinical Anatomy, Friedrich-Alexander-University Erlangen-Nuremberg, Erlangen, Germany
[2]Institute of Pathology, University Hospital Erlangen, Friedrich-Alexander-University Erlangen-Nuremberg, Erlangen, Germany
[3]Department of Otorhinolaryngology – Head & Neck Surgery, University Hospital Erlangen, Friedrich-Alexander-University, Erlangen-Nuremberg, Erlangen, Germany

***Keywords*: Surfactant proteins, Vocal folds, Human larynx, Molecular biology**

Abstract:

Objectives / Introduction: Surfactant proteins (SP) well known from lung have meanwhile been detected in other parts of the human body, e.g., in the larynx (Sheats et al., 2016). SP-G and PLUNC are two novel surfactant proteins that both have been recently detected in several tissues. SP-G (gene name: SFTA2) has so far been detected in e. g. human lung, eyelid, conjunctiva, lacrimal gland, kidney, and testis (Rausch et al., 2012). PLUNC has been detected in the epithelium and glands of the upper airways as well as in the nasal, tonsil and tongue glands (Bingle et al., 2005) and in the tissue of the tear system (Schicht et al., 2015). Until now, their functions are not yet fully known. Beside their surface activity they have been described to play important roles in the innate immune system (Bingle & Craven, 2002; Krause et al., 2019). This study describes the detection, characterization, and potential role of novel surfactant proteins SP-G and PLUNC in the human larynx with a focus on the vocal folds.

Methods: Expression and distribution of SP-G and PLUNC were analyzed by immunohistochemistry in healthy human tissue samples. Expression of both proteins was analyzed by Western blot in micro-dissected vocal fold mucosa. An immortalized human hypopharyngeal cell line (FaDu HTB-43) was used as an *in-vitro* model of the vocal fold epithelium. Cells were stimulated with scratch (mechanical damage) as well as with cortisol, serotonin and bombesin, followed by quantification of mRNA-expression of SPs using qRT-PCR. Furthermore, the possible role of the SP-G and PLUNC during wound healing was investigated by means of in vitro wound healing assay using ECIS.

Results: Presence of SP-G and PLUNC was demonstrated in laryngeal epithelial cells and glands. SP-G was found in squamous cell carcinoma of the vocal folds, whilst PLUNC was not. Mechanical stress (scratch assay) and stimulation with cortisol and serotonin induced the expression of SP-G and PLUNC in the FaDu cells. The addition of SP-G and PLUNC led to an accelerated wound healing of FaDu cells.

Conclusions: The results demonstrated that SP-G and PLUNC are surfactant proteins of the human larynx that have potential wound healing properties. Presence of SP-G was found to be related to vocal fold squamous cell carcinoma. Based on our results, we assume that the proteins may play an important role during wound healing and tumor pathogenesis in the human larynx respectively the vocal fold.

Acknowledgements:

The authors sincerely thank those who donated their bodies to science so that anatomical research could be performed. Results from such research can potentially increase knowledge that can improve patient care. Therefore, these donors and their families deserve our highest gratitude. Moreover, the authors would like to thank the patients for donating samples for our research. The authors would also like to thank Anke Fischer-Größwein, Maike Hemmerlein, , Hong Nguyen, Jörg Pekarsky and Lisa Stache for their technical support.

LRIG1 CELLS AS EPITHELIAL STEM CELLS FOR VOCAL FOLD MAITENANCE AND REGENERATION

Vlasta Lungova, Jessica M Fernandez, Susan L Thibeault
Department of Surgery, University of Wisconsin Madison, Madison, WI, USA

***Keywords*: Vocal folds; Epithelial stem cells; Genetic manipulation; Lineage tracing**

Abstract:

Objectives / Introduction:

Dysfunction of vocal fold (VF) epithelium may play a significant role in the etiology of VF benign hyperplastic lesions. Typical pathological alternations include squamous hyperplasia, mild, moderate and severe dysplasia (Gürbüz *et al.*, 2013). Hyperplastic lesions can cause the irregularities in the VF vibrations as they are located on the free edge(s) of vibrating portions of the VFs. They can increase the regional mass, impeding complete glottic closure, diminishing vibration and causing hoarseness and poor voice quality (Gürbüz *et al.*, 2013).

In studies of the epithelium from elsewhere in the body, cell types that are responsible for hyperplasic changes are adult epithelial stem cells (EpSCs). These cells can self-renew and give rise to the progeny of differentiated daughter cells (Byrd *et al.*, 2019). In stratified epithelia, EpSCs reside in the basal cell layers, retain a BrdU label and exclude Hoechst dye, which are nonspecific methods for EpSC identification, nowadays replaced by more precise genetic labeling targeting the Lrig1 gene (Byrd *et al.*, 2019). Lrig1 labels quiescent and actively dividing EpSC that participate in homeostasis, and tissue repair. Due to their plasticity and differentiation capabilities, these cells can serve as excellent candidates for genetic manipulation allowing for the modeling epithelial hyperplasia both *in vivo* and *in vitro* conditions (Succony *et al.*, 2016)

The main aim objective of this study was to elucidate the distribution of Lrig1-expressing EpSCs within the murine and human VF epithelium. Additionally, we aimed to assess the differentiation potential of Lrig1 cells via genetic approaches, employing lineage tracing in a murine model.

Methods:

To genetically label Lrig1 cells with tdTomato Red Fluorescent Protein (tdTom), we generated Lrig1-CreERT2 R26tdTom mice. Activation of tdTom and induction of recombination were achieved by administering tamoxifen (TX) to adult Lrig1-CreERT2 R26tdTom mice. For lineage tracing, two consecutive TX injections were applied and larynges were harvested 1, 5day(s), 1week and 1month post-recombination (PR). Larynges were immediately fixed, processed and embedded in Optimal Cutting Temperature Compound (OCT) for cryosectioning and used for histological analyses, such as immunofluorescent (IF) staining. To confirm presence and quantity of tdTom (Lrig1) cells in murine larynges, we sorted the cells. Larynges were harvested 1day PR and were enzymatically dissociated to form a single cell suspension and then sorted to separate tdTom+(Lrig1+) from tdTom- (Lrig1-) cells. Tdtom R26 mice were used as controls in these experiments. Human larynges were utilized from our repository. Adult larynges were fixed in 10% formalin, VF dissected, processed, and embedded in paraffin for sectioning and histological analyses (Hematoxylin & Eosin and IF staining). Similarly, fetal human larynges were fixed in 4% paraformaldehyde, processed, and paraffinembedded for sectioning and histological analyses, as mentioned above. To visualize Lrig1 cells in both adult and fetal human VF epithelium, paraffin-embedded sections were stained by an anti-Lrig1 antibody following our established protocols.

Results:

Our data revealed the presence of tdTom expressing Lrig1 cells in the VF epithelium, in mice, persisting from 1 day up to 1 month PR (Fig. 1A-H). These cells were located in the basal cellular compartment and transitioned into suprabasal layers as they differentiated. Notably, in the midmembraneous portion of VF, we observed a faster turnover of tdTom (Lrig1) cells compared to those in the supraglottic (cranial) and subglottic (caudal) VF regions suggesting an increased abrasion and functional demands of this specific region of the VFs. IF staining provided additional confirmation that tdTom (Lrig1) were located in the basal cell layer (co-localization with p63) and capable of cell proliferation and stratification (co-localization with Ki67 and cytokeratin 14) (Fig. 1I-K). Flow cytometry cell sorting yielded a total of 4.5% cells which aligns with the quantity of stem cells obtained from the epithelial tissue (Yamashita *et al.*, 2007) (Fig. 1L). We also detected Lrig1-expressing cells in human adult and fetal VF tissues. While in adult VF epithelium, Lrig1 cells exhibited scattered distribution, in fetal tissues the Lrig1 signal was diffused through the entire epithelium (Fig. 1M-P).

Conclusions:

We have successfully identified and localized Lrig1-expressing cells in murine VF epithelium and monitored their behavior for a period of one month. Similarly, we detected Lrig1 cells in human adult and fetal VF tissues. Given the Lrig1 cell position and notable proliferation/differentiation potential, these cells are potentially EpSCs capable of instigating hyperplastic changes in the epithelium. Further investigation is necessary to elucidate the Lrig1 cell specific requirements for selfrenewal and differentiation under normal conditions and in response to the stress factors. Additionally, conducting single cell RNA sequencing from the Lrig1 cell population will be beneficial to confirm their progenitor/stemness phenotype and capabilities to induce hyperplastic alternations in VF epithelium.

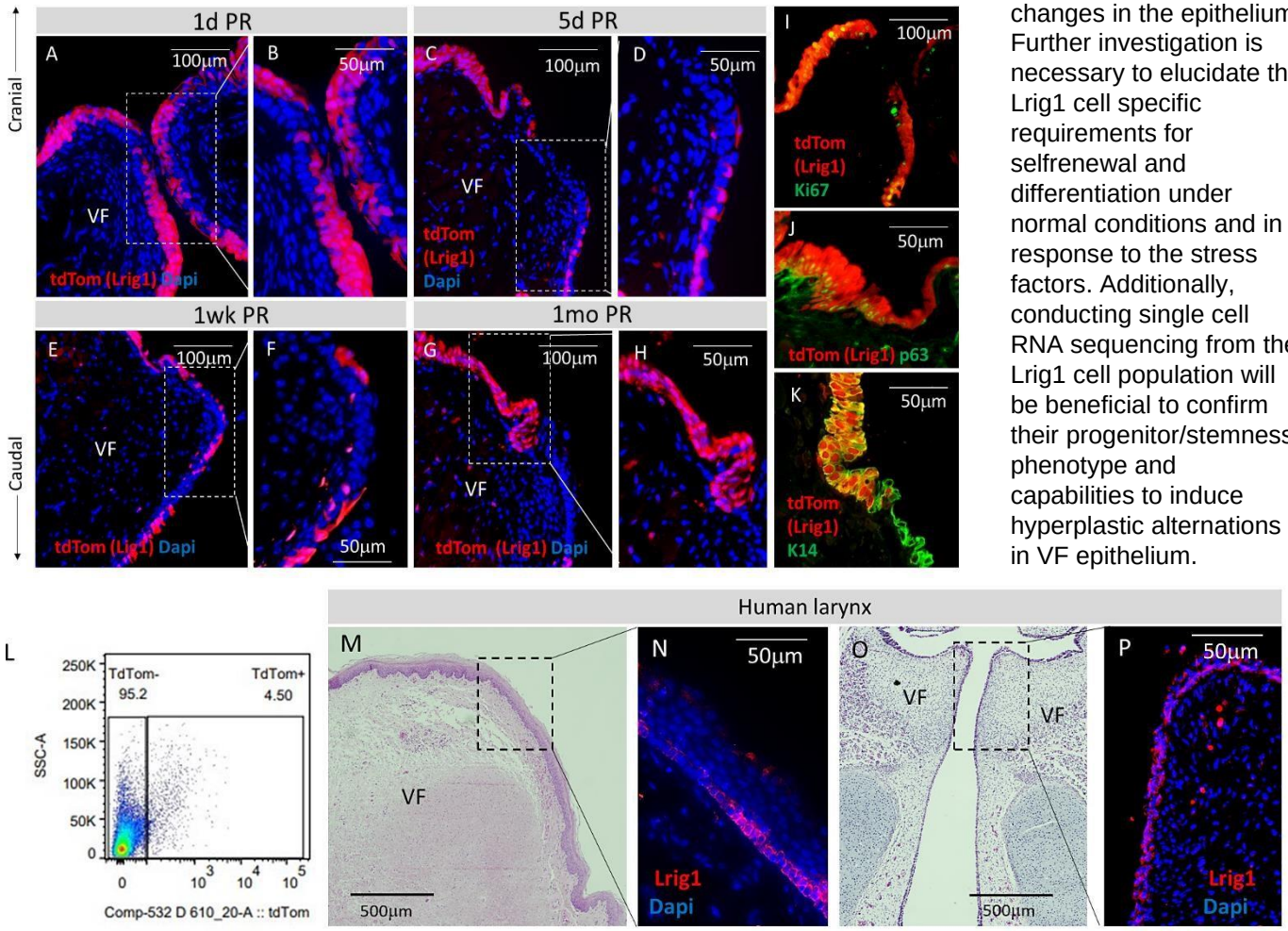


Figure 1: *Lineage tracing of tdTom labelled Lrig1 cells in a murine model (A-H). TdTom co-expression in red with cell proliferation marker Ki67 in green (I), basal cell marker p63 in green (J) and stratification marker cytokeratin 14 in green (K). Flow cytometry cell sorting for tdTom labelled cells (L). Hematoxylin & eosin and Immunofluorescent antiLrig1 staining in human adult (M,N) and fetal (O,P) vocal folds.. Abbreviations: VF, vocal folds.*

Acknowledgements:

This study was funded by NIDCD R01004336, R01012773, R01020734 and UW Madison Sophomore Scholarship.

References:

Byrd, K. M. *et al.* (2019) “Heterogeneity within Stratified Epithelial Stem Cell Populations Maintains the Oral Mucosa in Response to Physiological Stress,” *Cell Stem Cell*, 25(6), pp. 814-829.e6. doi: 10.1016/j.stem.2019.11.005.

Gürbüz, M. K. *et al.* (2013) “Vocal fold hyperplastic lesions: An evaluation of surgical outcome with videolaryngostroboscopy,” *Balkan Medical Journal*, 30(2), pp. 172–177. doi: 10.5152/balkanmedj.2012.113.

Succony, L. *et al.* (2016) “Role of LRIG1-dependent EGFR signalling on pathway inhibition in airway homoeostasis and lung cancer development,” *The Lancet*. Elsevier Ltd, 387, p. S95. doi: 10.1016/s01406736(16)00482-7.

Yamashita, M. *et al.* (2007) “Side population cells in the human vocal fold,” *Annals of Otology, Rhinology and Laryngology*, 116(11), pp. 847–852. doi: 10.1177/000348940711601110.

PARAMETER OPTIMIZATION OF AGENT-BASED MODELS FOR VOCAL FOLD BIOMATERIAL DESIGN

Grace Yu[1], Michael Döllinger[2], Nicole Y. K. Li-Jessen[3]

[1] Mayo Clinic Medical Scientist Training Program, Mayo Clinic, Rochester, Minnesota, USA

[2] Division of Phoniatrics and Paediatric Audiology, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

[3] School of Communication Sciences and Disorders, McGill University, Montreal, Quebec, Canada

***Keywords*: Vocal Folds, Biomaterials, Tissue Engineering, Computational Modeling**

Objectives / Introduction

Regenerative biomaterials offer promising therapeutic potentials for severe vocal fold (VF) defects, such as surgical scars and atrophy. However, biomaterial design requires the tuning of numerous chemical and physical parameters, where slight variations can change the subsequent biological responses of cells embedded in the biomaterial or native to the host tissue[1]. Computational models could be leveraged to guide and accelerate tissue-specific biomaterial design.

Agent-based models (ABMs) are mechanistic computational models that are comprised of a collection of autonomous entities, i.e. agents, that follow a set of pre-determined behaviors, i.e. rules, to simulate macroscopic, emergent phenomena. For example, ABMs can simulate individual cells as agents, generating collective temporal and spatial dynamics of cells and extracellular matrix remodeling at the tissue- or organ-level.

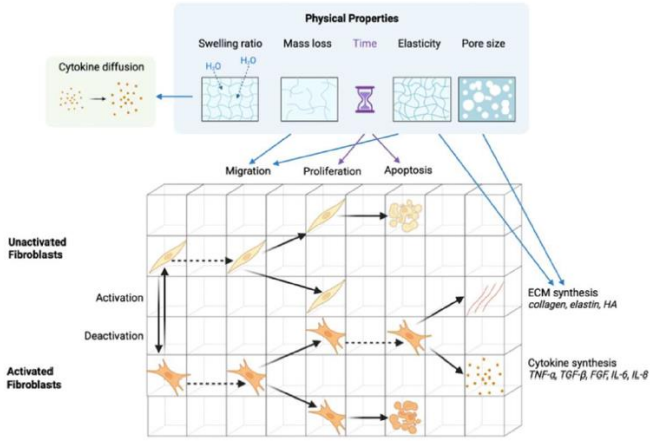


Figure 1. Schematic of VFB-ABM agent rules.

We developed an **agent-based model of VF biomaterials (VFB-ABM)** that simulates cell-material and cell-cell interactions in hyaluronan-based hydrogels for vocal fold biomaterial design **(Figure 1)**[2]. High-fidelity simulations from the VFB-ABM require many parameters with uncertain values, so parameter optimization is crucial to improving the accuracy of the model. Further, the stochastic, non-linear, and high-dimensional nature of the VFB-ABM compound the challenge for efficient and effective model calibration[3].

Previously, we tested four unconstrained multivariate numerical optimization methods: **Nelder-Mead (NM), BroydenFletcher-Goldfarb-Shanno (BFGS), conjugate gradient (CG),** and **Powell**. These methods were selected for their variety in algorithms. NM is a gradient-free simplex method. BFGS and CG are gradient-based methods. Powell is a gradient-free line search method. We found that NM was most effective and efficient. That being said, beyond optimization techniques, a more suitable objective function can further improve the optimization of the VFB-ABM, thereby improving its accuracy.

Hence, in this study, we aimed to improve the accuracy of the VFB-ABM by tuning the objective function used in the optimization algorithms. Specifically, we examined the effect of including additional timepoints or model parameters in the objective function on the efficiency and efficacy of VFB-ABM optimization.

Methods

The VFB-ABM was written in C++. The simulated biomaterial world was set as 0.6 x 0.6 x 0.6 mm^3 and initially seeded with 125 fibroblast cell agents. Model inputs were material compositions including hyaluronic acid, gelatin, and PEGDA crosslinker concentrations. Outputs included physical properties of materials (swelling ratio, mass loss, elasticity, and pore size), inactivated and activated fibroblast counts, as well as concentrations of cytokines (TNF-□, TGF-□, FGF, IL-6, and IL-8) and extracellular matrix proteins (collagen, hyaluronic acid, and elastin).

Sensitivity analysis was first performed on the 75 agent-rule parameters, whose values were not reported in literature. The sensitivity analysis techniques Morris, a statistical method, and Random Forest, a machine learning method, were tested and compared. Subsequently, parameter optimization was performed on the three or five most influential parameters, as determined by sensitivity analysis. The objective function for each day is defined in Equation 1, where E is error, F is the number of fibroblasts, C is the amount of collagen, and the subscripts s and e refer to simulated and experimental outputs respectively.

$$E = (F - F) + (C - C) \quad \textbf{Equation 1}$$

The overall objective function was given by the sum of objective functions for each of the days included: either day 9 only, or days 3, 6, and 9. The relative decrease in error (%) given by the objective function and efficiency, given by number of VFB-ABM iterations and computational time required, were evaluated.

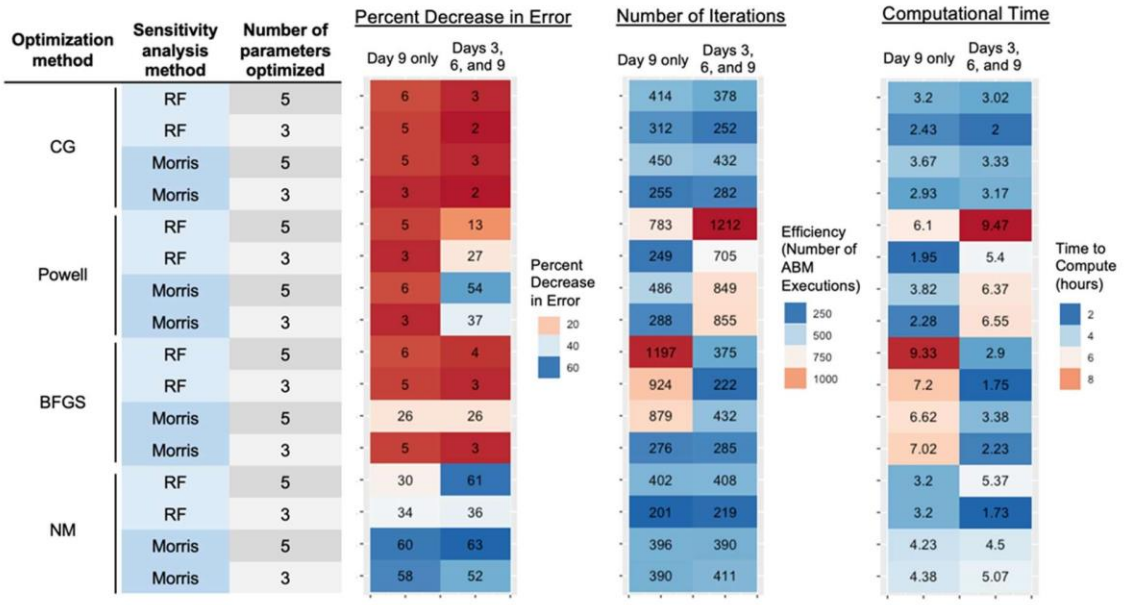


Figure 2. VFB-ABM optimization results for different objective functions.

Results

We found that the use of more parameters (n = 5) improved efficacy of the optimization, with a relatively small increase in computational cost (Figure 2). Also, optimization that was performed over all three timepoints was more effective for Powell and NM, but comparable for CG and BFGS, with mixed effects on efficiency depending on the optimization technique. Overall, using NM to optimize the top 5 parameters from Morris sensitivity analysis over all timepoints produced the best VFB-ABM parameter optimization results, yielding a decrease in the absolute error from 4.04 to 2.43, or a 63% relative decrease in error, and taking 390 ABM iterations or 4.5 hours to execute.

Conclusions

NM optimization enabled an effective parameter search and efficient narrowing of the parameter space to minimize the VFB-ABM error. Including more experimental timepoints and parameters in the objective function improved optimization accuracy with modest tradeoffs in efficiency. The VFB-ABM is currently limited to only hyaluronic acidgelatin biomaterials seeded with fibroblasts, and in vitro simulations. We are currently working on docking the VFBABM to an ABM of vocal fold tissue[4] to simulate in vivo responses to biomaterials. The ultimate goal is to leverage computer models to identify the most promising biomaterial designs and reduce the need for animal studies.

Acknowledgements

This project is supported by the National Sciences and Engineering Research Council of Canada (RGPIN-2018– 03843 and ALLRP 548623-19), Compute Canada and Canada Research Chair research stipend (N.L.J.). 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. Coburn PT et al. Journal of Biomedical Materials Research Part A 109(8) (2021) 1337-1352.
2. Shung C. McGill University, 2017.
3. Granato B, et al. ECAI 2020, IOS Press2020, pp. 2905-2906.
4. Garg A, et al. Appl Sci 9(15) (2019) 2974.

PERCEPTION OF VOICING IN SPEECH PRODUCED BY ADULT AND PEDIATRIC KINEMATIC VOCAL FOLD MODELS

Elizabeth Heller Murray[1], Brad Story[2], Terra Baptiste[1], Anna Rutherford[1]

[1] Department of Communication Sciences and Disorders, Temple University, Philadelphia, PA, USA
[2] Department of Speech, Language, and Hearing Sciences, University of Arizona, Tucson, AZ, USA

***Keywords*: Voice; Perceptual Cues for Voicing; Kinematic Vocal Fold Model**

Objectives / Introduction:

Production of stop consonants requires coordination of the articulatory and vocal mechanisms. During voiceless productions (p,t,k), the vocal folds typically cease vibrating during the consonant and restart for the vowel. Conversely, vocal fold vibration does not need to stop during the production of voiced stop consonants (b,d,g). This study aimed to examine the effect of maximum abduction width on the perception of voicing in stop consonants in speech produced by an adult and pediatric kinematic vocal fold model.

Methods:

Stimuli: Stimuli were generated as vowel-consonant-vowel (VCV) utterances using a kinematic model of the vocal folds coupled to a parametrically controlled vocal tract model where the consonant was either bilabial or alveolar. The consonant can be made voiceless by temporal control of abduction occurring in parallel with timing of the vocal tract occlusion. To generate a continuum from a voiced to voiceless consonant, separation of the vocal processes at the maximum point of abduction was increased from 0 cm^2 to 0.1 cm^2 in steps of 0.01 cm^2 in a collection of three versions of abductory timing functions for which the duration of the maximum abduction was varied. The vocal tract and abductory timing modulations were used in the context of an adult male and 4-year-old child speech production systems. A total of 132 unique stimuli were created for this study: 2 speakers (adult, child) * 2 places of articulation (bilabial, alveolar) * 11 points on the abduction continuum * 3 timing conditions = 132 stimuli.

Experimental design: Data was collected from 30 female participants who spoke American English as their primary language and passed a hearing screening. Participants were seated in a sound booth and heard stimuli through headphones at 75 dB. The experimental paradigm was a 2-alternative forced choice task, in which participants indicated whether they heard a voiced or voiceless. Presentation order was blocked so participants either heard the adult speaker first or the child speaker first. Participants were randomly presented with four repetitions of each stimulus within each block.

Results:

The percent of trials identified as voiced were calculated for each participant at each point on the continuum for each condition. A probit model was fit to the percent trials identified as voiced, and an inverse prediction function was applied to identify the 50% point on the stimuli continuum. Consistent with previous work, this point on the curve was interpreted as a boundary indicating a perceptual shift from voiced to voiceless. An
Analysis of Variance (ANOVA) found a significant effect of speaker, place of articulation, and timing condition

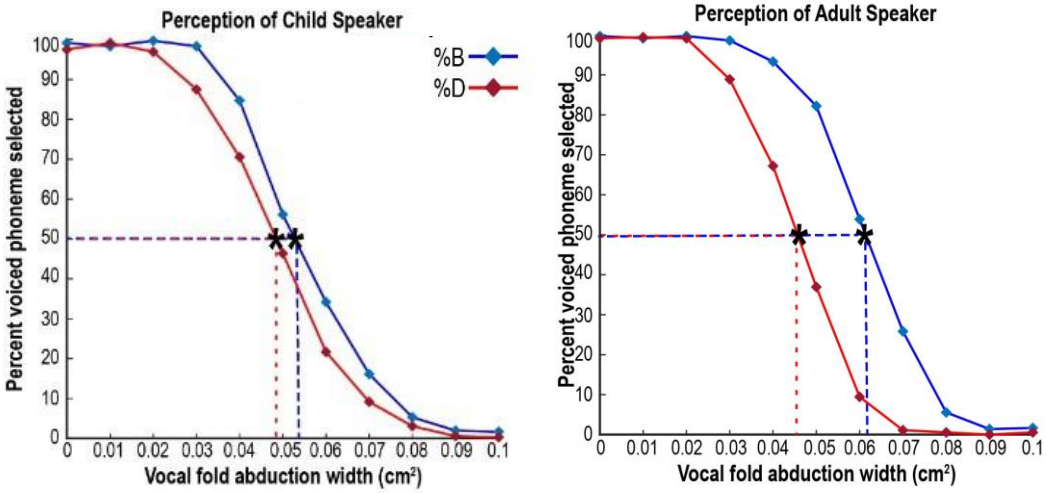


Figure 1. Response curves for the child speaker (left) and the adult speaker (right) indicating the percent of voiced phonemes identified at each vocal fold abduction width. The 50% point on each curve are indicated with an asterisk for the bilabial (blue) and alveolar (red) curves .

on the 50% point identified on the probit curve (F = 27.88, $p < 0.001$). There was a significant effect of condition ($p < 0.001$). The shift from voiced to voiceless was identified at 0.06 cm^2 for the condition in which maximum abduction was held for the shortest amount of time. During the other two conditions, the shift from voiced to voiceless was identified at a significantly smaller abduction width (both conditions, M = 0.05 cm^2). Additionally, the ANOVA indicated there was a significant interaction of speaker and place of articulation ($p < 0.001$, Figure 1). To examine the interaction further, the 50% point identified on the alveolar continuum was subtracted from the 50% point identified on the bilabial continuum. There was a significantly larger difference when perceiving the adult speaker (M = 0.016 cm^2) as compared to when perceiving the child speaker (M = 0.006 cm^2). This difference was primarily driven by the bilabial condition, as the average 50% point for the adult speaker was 0.061 cm^2 compared to the average 50% point for the child speaker at 0.055 cm^2. The average 50% point for the alveolar condition for the adult and child were similar, at 0.045 cm^2 and 0.048 cm^2.

Conclusions:

In the shortest timing condition, participants required a greater abductory width before they perceived a production as voiceless compared to the two longer timing conditions. There was no significant difference between the two longer timing conditions, suggesting that the timing of the abduction may only contribute to the perception of voicing if the maximum abduction width is held for a short amount of time. There was also a significant interaction between the speaker and the place of articulation. In the perception of the child speaker, listeners perceived a shift from the voiced consonants (b, d) to the voiceless consonants (p, t) at comparable abductory widths. Perception of a voicing shift in the adult speaker yielded similar results for the alveolar continuum (d to t) as the perception of the child speaker. Yet, a significantly larger abductory width was required to perceive a voicing shift in the bilabial continuum (b to p). One potential explanation for the difference in the perception of adult and child speech may be related to the listeners' expectations of speech production. Children's vocal and articulatory motor control are less exact and coordinated than adults, with developmental improvements in motor control resulting in increased intelligibility during development. Therefore, if listeners anticipate reduced intelligibility or less exact speech in children, widening the vocal fold abduction width may provide sufficient acoustic cues to differentiate voice and voiceless productions. Alternatively, when perceiving the speech of a mature adult speaker, listeners may anticipate more subtle and complex cues that are secondary to abduction width, such as cues from the burst or formant transitions. This may be even more applicable when observing the bilabial (b-p) production, as the short vocal tract during this production may make it more difficult to perceive a voiceless consonant without additional cues, such as changes in burst amplitude.

Acoustic analysis is ongoing, but preliminary results suggest that one of the primary cues for voicing contrasts, voice onset time (the time between the burst and the start of the vocal fold movement), alone does not explain these perceptual differences. In fact, using this kinematic model, voice onset time was primarily 0 milliseconds, meaning the vocal folds were consistently vibrating throughout the consonants. The widening of the vocal fold abduction only resulted in ceasing the vocal fold vibration at the largest few abduction widths. Thus, when listening to both adults and children, listeners would need to rely on other cues to determine voicing categories. Preliminary work on the adult stimuli suggests that first formant transitions may be particularly important in determining the voicing category. Figure 2 shows trajectories for one of the timing conditions, with the formant trajectory plotted over abduction widths ranging from 0.03 cm^2 to 0.09 cm^2. The asterisk in the figure marks the approximate abduction width in which participants shifted from a perception of voiced to voiceless, with the shift for the bilabial (b-p) continuum occurring at a larger abduction width than the alveolar (d-t) continuum. The dotted lines indicate places where formant values are less stable, frequently masked by aspiration noise from the consonant. Initial examination suggests that a key indicator of voicing status is the timing of the first formant stability, also called F1 cutback. There is sometimes a less prominent F1 transition in voiceless productions, as the articulatory gesture is primarily complete when the voicing begins. The final analysis will include a discussion of the impact of additional acoustic cues on the perceptual findings in the results section.

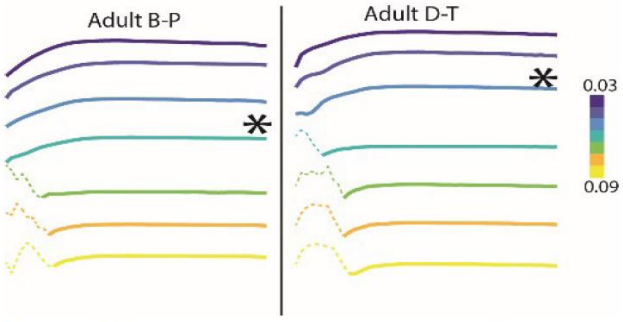


Figure 2. First formant trajectories for vocal fold abduction width 0.03 – 0.09 cm^2. Asterisk marks point around the 50% mark on each of the curve. Dotted lines in the formant traces are places where the formant tracking was less stable.

Acknowledgements: Thanks to Bella Pellitta for help with data collection.

MODELLING THE INFLUENCE OF LOCALIZED EDEMA ON VOCAL FOLD KINETICS AND KINEMATICS

Jonathan J. Deng[1], Byron D. Erath[2], Matías Zañartu[3], Sean D. Peterson[1]

[1] Department of Mechanical and Mechatronics Engineering, University of Waterloo, Waterloo, Ontario, Canada
[2] Department of Mechanical and Aerospace Engineering, Clarkson University, Potsdam, New York, U.S.A
[3] Department of Electronic Engineering, Universidad Técnica Federico Santa María, Valparaíso, Chile

***Keywords*: Finite element modelling, Vocal hyperfunction, Nodules, Swelling, Edema**

Abstract:

Objectives / Introduction: Vocal hyperfunction (VH) is a class of prevalent voice disorders with complex and poorly understood etiology. Phonotraumatic vocal hyperfunction (PVH), a subclass of VH, is characterized by excessive collision stresses and is implicated in the formation of benign vocal fold (VF) lesions, such as nodules, through a feedback loop termed the "vicious cycle". This feedback loop is believed to be initiated by some trigger, which then results in a change in voice, such as loudness or fundamental frequency. The speaker then compensates in order to return their voice features back to the normal range. The new baseline results in more forceful phonation, which further exacerbates the deleterious effects of the initial trigger and the cycle progresses, potentially resulting in a phonotraumatic lesion.

Swelling (edema) of the VFs is one hypothesized trigger of the vicious cycle, though the relationship between swelling and alterations in voice outputs and measures of tissue damage is unclear. In our previous work, the effect of swelling on measures of phonotrauma and voice outputs was investigated for increasing degrees of systemic swelling in the cover layer of a two-dimensional finite element (FE) vocal fold model [Deng2023]. Swelling was found to elevate phonotrauma-related measures, including collision pressure, viscous dissipation (associated with dissipation dose [Titze2003]), and von Mises stress (associated with distortion energy [Hill1998]), while simultaneously reducing fundamental frequency. The reduction in fundamental frequency can lead to heightened subglottal pressure, increased intrinsic muscle activation, or the initial of other compensatory responses. Given these effects, along with the impact on phonotrauma indicators, we posited that swelling could play a role in the vicious cycle.

There were several limitations in this previous work; most notably, a two-dimensional (2D) model was employed with generalized swelling in the cover. Locally concentrated swelling is thought to be a precursor to the development of nodules [Johns2003]. These localized swelling regions generally occur in the mid-membranous portion of the VFs, thus necessitating a three-dimensional (3D) treatment of the problem. Secondly, the vicious cycle was not modeled, wherein swelling or other structural changes in the tissue are correlated directly with mechanical damage and may dissipate with healing. Modelling such components is important to predict whether a vicious cycle would occur and how quickly it might progress.

As an initial step towards addressing these limitations, we developed a 3D VF model to capture localized swelling based on spatial distributions of phonotrauma measures. We use this model to investigate how localized 3D swelling affects measures of phonotrauma and voice outputs to gauge the role of swelling in the vicious cycle.

Methods: A 3D FE VF model coupled with a 1D Bernoulli-based glottal flow solver was used to model VF vibration and measures of phonotrauma. The VF geometry, shown in Figure 1, employs the M5 geometry [Scherer2001] with a medial surface angle of 3^o, VF length of $1.5\ \mathrm{cm}$, and VF thickness of $0.7\ \mathrm{cm}$. The mesh comprised 4230 tetrahedral elements. For each coronal plane, 1D glottal flow is assumed and used to compute fluid pressures along the corresponding medial surface in that coronal plane. The VFs are assumed to be symmetric so that only a single VF was modelled. As in our previous work [Deng2023], swelling was modelled using a modification of a Saint-Venant Kirchoff hyperelastic material to favor a swollen volume [Tsai2004], while viscous effects were modelled through a Kelvin-Voigt model. As a first approximation, and unlike the previous 2D study, material stiffness is not impacted by swelling in the present study.

To model the effects of localized swelling on voice outputs and phonotrauma, self-oscillation of the VFs in an unswollen state was first simulated. Two different measures related to phonotrauma, viscous dissipation and elastic strain energy (a measure previously used to characterize damage in collagenous soft tissues [Balzani2012]), were

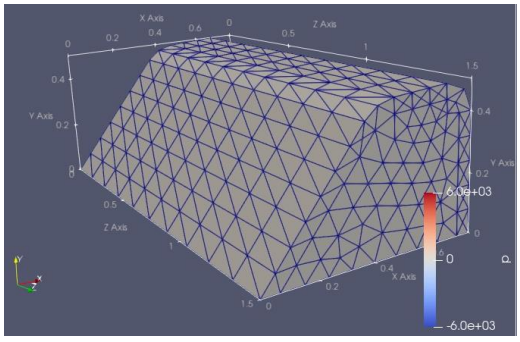


Figure 3: Vocal fold geometry and mesh.

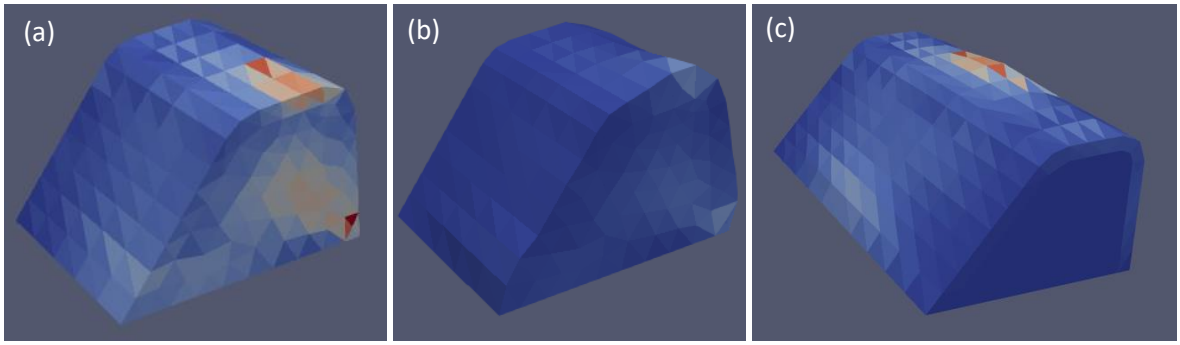


Figure 4: Time-averaged viscous dissipation over the last 0.25s of a 0.5s simulation with a slice through the mid-coronal plane for the (a) unswollen fold and (b) fold with maximum 15% swelling mapped linearly to viscous dissipation. (c) The resulting swollen fold colored by degree of swelling.

computed and averaged over time to obtain spatial distributions of phonotrauma. These time-averaged spatial distributions of phonotrauma were then used to prescribe localized swelling in the VFs via a linear mapping. Self-oscillation of the VF was simulated for increasing levels of localized swelling, and changes in voice outputs (fundamental frequency and sound pressure level) and changes in the two measures related to phonotrauma were computed. This was conducted for a range of total volume changes ranging from 1 to 1.3 times the unswollen volume.

Results: The two measures of time-averaged phonotrauma for the unswollen VFs show that damage is localized to specific regions of the VFs, including near the mid-membranous portion where the VFs contact (see the mid-coronal slice in Figure 2(a)). For viscous dissipation there is also a slightly elevated region in the body of the VF, and a highly localized concentration near the basement membrane along the superior margin of the fold. The portion along the medial surface shows relatively high values over a broad area, yet confined primarily to the cover where nodules are known to form [Titze1994].

Localized swelling distributed proportionally to viscous dissipation is applied to the VFs, as shown in Figure 2(c) with maximum 15% swelling. As a result, there is a noticeable change in the vibrational characteristics, with SPL decreasing by approximately 5 dB. Interestingly, and contrary to the 2D case studied previously [Deng2023], fundamental frequency remains largely unchanged despite the increase in VF mass, though this may not be the case when swelling-induced softening of the material is considered. In the absence of compensation to increase SPL back to the original level, the swollen fold exhibits reduced viscous dissipation, Figure 2(b), suggesting the potential for a protective measure, though the phonotrauma measure is still above base values and may result in further, albeit slight, swelling. We note that this "single step" model introduced significant swelling, so progression of the "vicious cycle" remains to be considered.

Conclusions: Phonotrauma-induced swelling was modeled using a 3D finite element VF model. Measures related to phonotrauma were found to localize in the cover layer, largely along the mid-membranous region of the VF. As a first attempt to model the vicious cycle, material swelling was linearly mapped to the magnitude of selected measures related to phonotrauma and the resulting VF dynamics were then simulated. Localized swelling was found to alter voice outputs by a sufficient margin to warrant compensation, such as increasing subglottal pressure, which may fuel the vicious cycle, though, in contrast to the 2D case, swelling was found to decrease viscous dissipation. Comprehensive exploration of this, considering compensation and multiple iterations, is planned for future research.

Acknowledgements:

Research reported in this work was supported in part by the National Institute on Deafness and Other Communication Disorders of the National Institutes of Health under award P50DC015446.

References:

Balzani, D. & Schmidt, T. (2012). Comparative analysis of damage functions for soft tissues: Properties at damage initialization. Mathematics and Mechanics of Solids, 20(4), 480-492.

Deng, J. J., Erath, B. D., Zañartu, M., & Peterson, S. D. (2023). The effect of swelling on vocal fold kinematics and dynamics. Biomechanics and Modeling in Mechanobiology, 22(6), 1873-1889.

Hill, R. (1998). The mathematical theory of plasticity (Vol. 11). Oxford university press.

Johns, M. M. (2003). Update on the etiology, diagnosis, and treatment of vocal fold nodules, polyps, and cysts. Current Opinion in Otolaryngology & Head and Neck Surgery, 11(6), 456-461.

Scherer, R. C., Shinwari, D., De Witt, K. J., Zhang, C., Kucinschi, B. R., & Afjeh, A. A. (2001). Intraglottal pressure profiles for a symmetric and oblique glottis with a divergence angle of 10 degrees. The Journal of the Acoustical Society of America, 109(4), 1616-1630.

Titze, I. R. (1994). Mechanical stress in phonation. Journal of Voice, 8, 99–105.

Titze, I. R., Svec, 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.

Tsai, H., Pence, T. J., & Kirkinis, E. (2004). Swelling induced finite strain flexure in a rectangular block of an isotropic elastic material. Journal of Elasticity, 75, 69-89.

EFFECTS OF FALSE AND ARYEPIGLOTTIC FOLD CONSTRICTIONS ON VOICE PRODUCTION IN A SIMPLIFIED VOCAL TRACT

Tsukasa Yoshinaga[1], Zhaoyan Zhang[2]

[1] Department of Mechanical Engineering and Bioengineering, Graduate School of Engineering Science, Osaka University, Toyonaka, Osaka, Japan

[2] Department of Head and Neck Surgery, University of Californica, Los Angeles, Los Angeles, Californica, USA

Keywords*: *Aeroacoustics; Source-filter Interaction; Fluid-structure Interaction; Two-mass Model

Abstract:

Objectives / Introduction: The voice is produced by the self-sustained oscillation of the vocal folds interacting with airflow from the lungs. While in general changes in vocal tract shape have small effects on the oscillation of the vocal folds, the effects can be significant under specific conditions [1]. For example, voice quality is known to be affected by adduction of the false vocal folds, which are located just above the true vocal folds and below the epiglottis. It has also been reported that constriction of the aryepiglottic folds may alter the efficiency of voice production [2]. Although there are many studies on this topic, e.g. [3], the effects of false and aryepiglottic fold constriction on voice production still remain unclear. In particular, since the aryepiglottic fold constriction restricts the airway in the anterior-posterior direction, it is expected to have strong aerodynamic effects on the vocal fold oscillations.

In this study, we perform a three-dimensional airflow analysis in a two-mass model coupled to a simplified vocal tract to examine how vocal tract constrictions at the levels of the false and aryepiglottic folds affect the oscillation characteristics and sound generation of the true vocal folds.

Methods: Vocal fold oscillation was simulated using a two-mass model, which has smoothly connected surfaces on two cylindrical masses [4]. The springs and dampers were connected to each mass to reproduce the mucosal waves. The true vocal fold model was placed in a 17 × 17 mm rectangular channel in between the sub- and supra-glottal tract with lengths of 150 and 175 mm, respectively. The thickness of the vocal folds was set to 3 mm (0.5 mm for the upper masses and 2.5 mm for the lower masses). Rigid false and aryepiglottic fold models were placed downstream of the true vocal folds as shown in Fig. 1. The false folds with a length of 4 mm were set parallel to the true folds at 4 mm downstream from the true folds. The aryepiglottic folds with a length of 8 mm were oriented along the medial-lateral direction at 12 mm downstream from the true folds. The shape of the constriction was designed using a Gaussian function similary as in [5]. Both the false and aryepiglottic folds were rigid, and deformation due to airflow pressures was not taken into account. The width of the constrictions was varied from 3 to 13 mm to investigate the effect of each constriction. At downstream from the vocal tract, we set a space of 230 × 160 × 160 mm^3 where the sound can propagate away from the vocal tract exit.

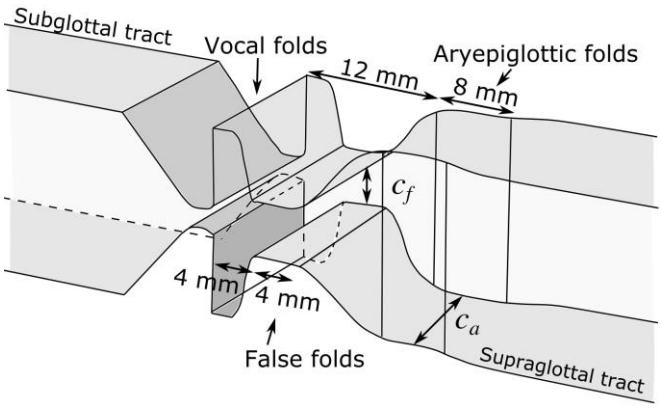


Fig. 1 Glottal geometry for the simulation.

In order to simulate the airflow and sound generated by the true vocal folds' oscillation, compressible Navier-Stokes equations were solved using the finite difference method as reported in [6]. The structured computational grids were constructed on the simplified vocal tract model described above. For the spatial differentiation, a 6th-order-accuracy compact scheme was used, while the time integration was calculated using the 3rd-order-accuracy Runge-Kutta method. The spatial filter with 10th-order accuracy was used as an implicit large eddy simulation to consider the turbulence of the airflow in the larynx. The moving fluid-structure interface on the structured grids was modeled using the volume penalization method. A pressure chamber was placed upstream from the subglottal tract, and a uniform flow rate of 350 cm^3/s with a constant pressure of 1200 Pa was set to reproduce the normal phonation condition in the model. The sound pressure levels (SPLs) were calculated by sampling a pressure fluctuation at 160 mm from the vocal tract outlet, where the flow velocity is small (< 0.02 m/s).

Results: The mean fundamental frequency *f*o, closed quotient (CQ), and SPL are plotted in Fig. 2. The *f*o increased by 9 Hz with the false fold constriction, whereas the *f*o decreased by 8 Hz with the aryepiglottic fold constriction. There was little change in CQ with changes in the constriction sizes from 5 to 13 mm for both constriction locations. When the constriction size was decreased from 5 to 3 mm, the CQ increased by 0.04 with the false fold constriction, while the CQ decreased by 0.02 with the aryepiglottic constriction, indicating a negligible effect on CQ. The comparison of outside sound pressures (Fig. 2(c)) showed that narrowing the false folds' gap from 13 to 3 mm increased the sound pressure by more than 5 dB, whereas the narrowing of the epiglottic folds changed the SPL only within ±1 dB and had smaller effects compared to the false folds.

A previous study using a one-dimensional glottal flow model [5] demonstrated that vocal tract adjustments including aryepiglottic constrictions have relatively small effects on the glottal source except for the extreme constriction conditions, and the current results are consistent with the previous study. However, we did not observe a significant increase in the far-field SPLs with the aryepiglottic fold narrowing, which is usually observed in human phonation [2]. This small increase was likely related to the specific vocal fold condition and vocal tract geometry used in this study [5], which will be clarified in our future studies.

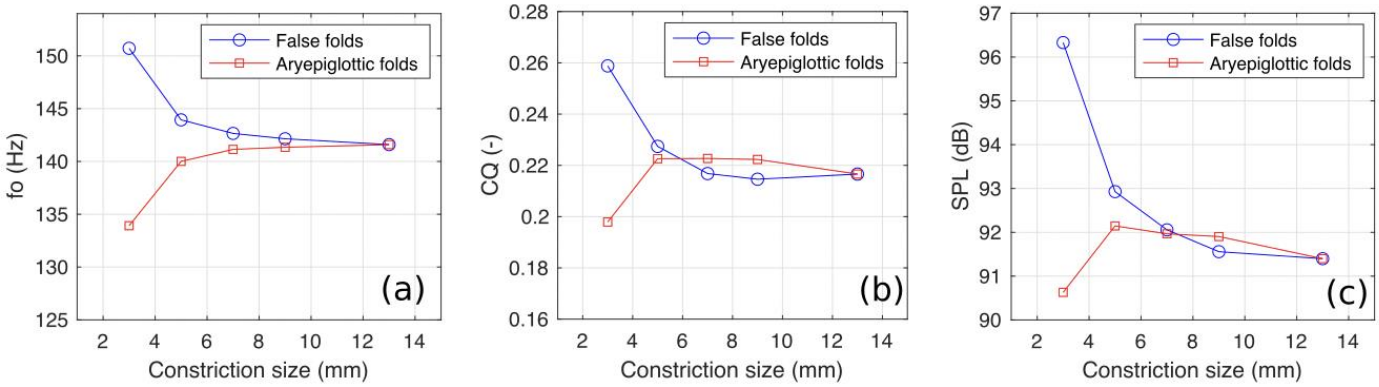


Fig. 2 Mean fundamental frequency (a), closed quotient (b), and sound pressure level (SPL) at 100 mm from the vocal tract (c) plotted with different false and aryepiglottic folds' constrictions.

Conclusions: In this study, a three-dimensional compressible flow simulation was conducted to investigate the effects of false and aryepiglottic folds on voice outcomes. The results showed that both false and aryepiglottic folds had relatively small effects on the voice source characteristics, while the far-field sound amplitude was notably increased with the false fold constriction. Future studies will focus on changes in the glottal airflow and sound source characteristics associated with false and aryepiglottic constriction.

Acknowledgements:

This work was supported by MEXT as JSPS KAKENHI (Grant No. JP22KK0238; JP23K17195), and Research Grant No. R01DC020240 from the National Institute on Deafness and Other Communication Disorders, the National Institutes of Health.

References:

[1] Titze, I. R. (2008). Nonlinear source–filter coupling in phonation: Theory. *The Journal of the Acoustical Society of America*, 123(5), 2733-2749.

[2] Zhang, Z. (2021). The physical aspects of vocal health. *Acoustics today*, *17*(3), 60.

[3] Zheng, X., Bielamowicz, S., Luo, H., & Mittal, R. (2009). A computational study of the effect of false vocal folds on glottal flow and vocal fold vibration during phonation. *Annals of biomedical engineering*, *37*, 625-642.

[4] Pelorson, X., Hirschberg, A., Van Hassel, R. R., Wijnands, A. P. J., & Auregan, Y. (1994). Theoretical and experimental study of quasisteady-flow separation within the glottis during phonation. Application to a modified two-mass model. *The Journal of the Acoustical Society of America*, *96*(6), 3416-3431.

[5] Zhang, Z. (2023). The influence of source-filter interaction on the voice source in a three-dimensional computational model of voice production. *The Journal of the Acoustical Society of America*, *154*(4), 2462-2475.

[6] Yoshinaga, T., Zhang, Z., & Iida, A. (2022). Comparison of one-dimensional and three-dimensional glottal flow models in left-right asymmetric vocal fold conditions. *The Journal of the Acoustical Society of America*, *152*(5), 2557-2569.

COMPUTATIONAL ANALYSIS OF VOCAL FOLD PPOSTURING IN ASYMMETRIC LARYNGEAL MUSCLE ACTIVATIONS

Mahdi Sangbori[1], Weili Jiang[1], Biao Geng[1], Xudong Zheng[1], Qian Xue[1]

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

***Keywords*: Vocal Fold Paralysis; Vocal Fold Posturing; Flow-Structure Interaction**

Abstract:

Introduction:

The human voice is produced through intricate neuromuscular contractions of the intrinsic laryngeal muscles (ILM), establishing the prephonatory posture of the vocal folds [1]. Deficiencies in either the recurrent laryngeal nerve (RLN), superior laryngeal nerve (SLN), or both can result in ILM malfunction, leading to laryngeal paresis and paralysis [2]. The conditions mostly occur on one side, resulting in asymmetric activation of ILMs, which can cause changes in the shape of the prephonatory posture, a significant contraction in pitch range, asymmetry in vocal fold vibratory phase, and acoustic aperiodicity. These factors ultimately contribute to an overall decline in vocal quality [3]. Hence, it is essential to comprehend the role of asymmetric muscle activation and its influence on the shape and vibrational patterns of the vocal folds.

Several studies have sought to investigate the impact of symmetric intrinsic laryngeal muscle activation on vocal fold prephonatory shape, vibration, flow, and acoustics [4][5][6]. Other investigations have delved into the role of laryngeal paresis and paralysis, along with asymmetric muscle activation, aiming to understand their effects on vibration patterns, acoustics, and perception through numerical and experimental methods [2][7][8]. However, the understanding remains uncomplete. In particular, a systematic understanding of the effect of asymmetric ILM activations on vocal fold dynamics is incomplete. In this article, we propose to use high-fidelity computer modeling technique to investigate 3D vocal fold posturing in symmetric and asymmetric muscle activations, focusing on the relationships among muscle activity, vocal fold tension, stiffness, and geometry.

Methods:

Geometry model. A three-dimensional model of human larynx was reconstructed from high-resolution MRI scans of a 74-year-old female (Fig. 1). The model takes into account both left and right sides of all laryngeal cartilages—the thyroid, arytenoid, and cricoid cartilages—as well as the intrinsic muscles (CT, TA, lateral cricoarytenoid (LCA), interarytenoid (IA), posterior cricoarytenoid (PCA) muscles). These muscles are segmented into bellies, comprising the vertical (CTv) and oblique components (CTo) of the CT, as well as the thyromuscularis (TAm) and thyrovocalis (TAv) of the TA. The activation of muscles on each side can be independently regulated, allowing for both symmetric and asymmetric muscle activations. The vocal fold is a layered structure, encompassing the cover layer, ligament, conus elasticus (CE), TA, thyroid epiglottis (TE) muscle, paraglottic space, and cricoid space.

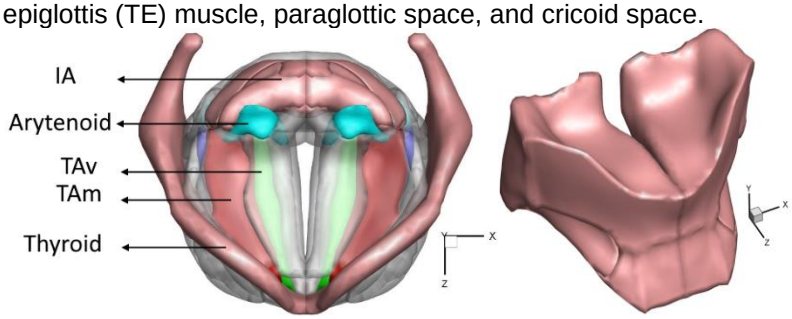


Figure 1. The geometric model of the human larynx reconstructed from the MRI scans of the larynx of a 74-year-old female. (Left) The whole larynx model for posturing simulation. (Right) The vocal fold model for FSI simulation.

Material property. A fiber-reinforced anisotropic material model [9] is employed to represent the passive material characteristics of the vocal fold cover, ligament, conus elastics and all the muscles. The model parameters for different tissues were optimized by aligning with experiment measurements. The active stress in muscles is modeled using the Hill-based contractile model. Neo-Hookean material is applied to cricoid, thyroid, and arytenoid cartilages, as well as the paraglottic space and cricoid space.

Simulation setup. Each simulation consists of two phases. In Phase 1, the posturing of the complete laryngeal model is simulated in response to muscle activations. The displacements of nodes on the anterior, posterior, and lateral surfaces of the vocal fold are extracted from the simulation results. These displacements serve as the boundary conditions for Phase 2. In this phase, an asymmetric activation ratio of the muscles is employed to represent both SLN and RLN paralysis conditions. Phase 2 exclusively simulates the flow-structure interaction (FSI) of the vocal fold, incorporating the displacement boundary conditions obtained in Phase 1. The medial, inferior, and superior surfaces of the vocal fold are allowed to move in response to aerodynamic loading. Tissue deformation is modeled using the finite

element method, and glottal flow dynamics is modeled using the one-dimensional Bernoulli equation. The three-dimensional geometries for both phases are illustrated in Fig. 1.

Results:

The changes in vocal fold geometry resulting from both asymmetric and symmetric ILM activations will be examined in detail. Fig. 2 shows preliminary results, including the variations in vocal fold length and vertical thickness in response to symmetric CT and TA activations. The baseline scenario represents a fully adducted condition with zero CT and TA muscle activations. The vocal fold strain resulting from the isolated activation of TA and CT muscles align with typical values reported in previous studies, suggesting that our simulation result is within the physiological range. CT activation leads to vocal fold thinning, while TA activation results in vocal fold thickening. This influence of CT-TA activation on vertical thickness is in line with the prior study in the canine model [4] and corresponds to experimental observations [10]. Asymmetric ILM activations will be conducted to understand their impacts on vocal fold pre-phonatory geometry and tension as well as vibration patterns. Fig. 3 shows two preliminary simulations of asymmetric activations. In both simulations, the CT, TA, IA, and LCA are fully activated on the left side. In Fig. 3a, the activation level of LCA and IA was reduced to 50% while keep other muscles the same as the left side. In Fig. 3b, the activation level of TA was reduced to 50% while keep other muscles the same as the left side. The results indicate a slight left-right asymmetry in vocal fold shapes.

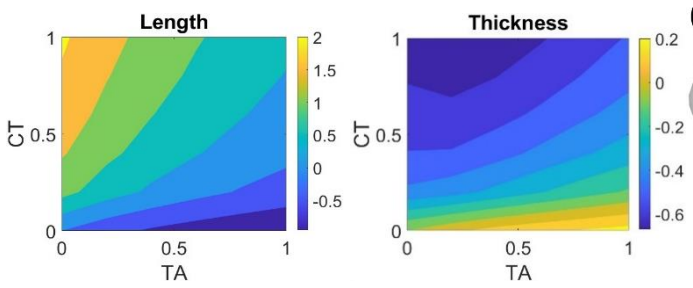


Figure 2. The variation of the vocal fold length and vertical thickness relative to the baseline condition in response to CT and TA activations.

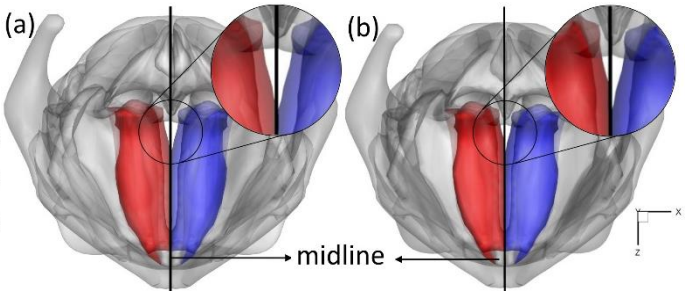


Figure 3. Asymmetric activation of (a) LCA/IA and (b) TA muscles

Conclusions:

A high-fidelity computer model of the human larynx, encompassing all cartilages and intrinsic laryngeal muscles on both sides, is reconstructed using MRI data. The model is used to quantify the 3D properties of vocal fold prephonatory posture during symmetric and asymmetric muscle activations representing SLN and RLN paralysis. The relationship between the asymmetric ratio of muscle activations and the asymmetric ratio of posture properties will be investigated. The FSI simulation is also employed to investigate the vibration patterns.

Acknowledgements:

The research was funded by NIH Grant No. 1R15DC019229.

References:

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

[2] D. K. Chhetri, J. Neubauer, and E. Sofer, "Influence of Asymmetric Recurrent Laryngeal Nerve Stimulation on Vibration, Acoustics, and Aerodynamics," *Laryngoscope*, vol. 124, no. 11, pp. 2544–2550, 2014.

[3] M. I. Orestes and D. K. Chhetri, "Superior laryngeal nerve injury: effects, clinical findings, prognosis, and management options," *Curr. Opin. Otolaryngol. Head Neck Surg.*, vol. 22, no. 6, p. 439, 2014.

[4] M. Movahhedi, B. Geng, Q. Xue, and X. Zheng, "Effects of cricothyroid and thyroarytenoid interaction on voice control: Muscle activity, vocal fold biomechanics, flow, and acoustics," *J. Acoust. Soc. Am.*, vol. 150, no. 1, pp. 29–42, Jul. 2021.

[5] B. Geng, N. Pham, Q. Xue, and X. Zheng, "A three-dimensional vocal fold posturing model based on muscle mechanics and magnetic resonance imaging of a canine larynx," *J. Acoust. Soc. Am.*, vol. 147, no. 4, pp. 2597–2608, Apr. 2020.

[6] A. M. Vahabzadeh-Hagh, Z. Zhang, and D. K. Chhetri, "Three-dimensional posture changes of the vocal fold from paired intrinsic laryngeal muscles," *Laryngoscope*, vol. 127, no. 3, pp. 656–664, Mar. 2017.

[7] H. R. Chung, Y. Lee, N. K. Reddy, Z. Zhang, and D. K. Chhetri, "Effects of Thyroarytenoid Activation Induced Vibratory Asymmetry on Voice Acoustics and Perception," *Laryngoscope*, 2023.

[8] T. Yoshinaga, Z. Zhang, and A. Iida, "Comparison of one-dimensional and three-dimensional glottal flow models in left-right asymmetric vocal fold conditions," *J. Acoust. Soc. Am*, vol. 152, pp. 2557–2569, 2022.

[9] N. Pham, Q. Xue, and X. Zheng, "Coupling between a fiber-reinforced model and a Hill-based contractile model for passive and active tissue properties of laryngeal muscles: A finite element study," *J. Acoust. Soc. Am*, vol. 144, pp. 248–253, 2018.

[10] A. M. Vahabzadeh-Hagh, Z. Zhang, and D. K. Chhetri, "Hirano's cover–body model and its unique laryngeal postures revisited," *Laryngoscope*, vol. 128, no. 6, pp. 1412–1418, Jun. 2018.

VOCAL3D: FULLY AUTOMATIC 3D RECONSTRUCTION OF PHONATING HUMAN VOCAL FOLDS VIA STRUCTURED LIGHT LARYNGOSCOPY

Jann-Ole Henningson[1], Marion Semmler[2], Reinhard Veltrup[2], Michael Döllinger[2], Marc Stamminger[1]

[1] Visual Computing, Friedrich-Alexander-Universität Erlangen-Nürnberg, 91058 Erlangen, Germany
[2] Division of Phoniatrics and Pediatric Audiology at the Department of Otorhinolaryngology, Head and Neck Surgery, University Hospital Erlangen, Friedrich-Alexander-Universität Erlangen-Nürnberg, 91054 Erlangen, Germany

***Keywords*: Structured Light; 3D Reconstruction; Image processing; Machine Learning**

Abstract:

A current branch of voice research is interested in the 3D reconstruction of human vocal folds during phonation. This allows not only to measure the lateral and longitudinal, but also the vertical deformation of human vocal folds. At the same time, a 3D reconstruction increases the plasticity of visualization and may aid physicians in better detecting malicious diseases. To achieve this, endoscopic structured light systems are currently being explored that project a known point pattern into the laryngeal area. One can then generate 3D world positions of the superior surface of the vocal folds by measuring the deformation of this point pattern. However, to achieve clinical applicability of 3D Laryngoscopy, it is of upmost importance that these reconstruction pipelines work robustly without the need of human intervention. For this, we present **Vocal3D**: A fully automatic 3D reconstruction pipeline for Structured Light Laryngoscopy. The source code is open source and can be found here: https://github.com/Henningson/Vocal3D

Method:

The pipeline can generally be considered to consist of four parts:

- A Feature Detection step, that segments glottal gap and vocal folds and detects laser point positions.
- Next, a correspondence estimation step, that estimates which laser beam created which observed laser point.
- Then, a triangulation step of the per frame point cloud in which the reconstructed point cloud is split based on the glottal midline.
- Lastly the 3D reconstruction step, in which we fit a B-Spline based M5 model into the point clouds, based on the As-Rigid-As-Possible deformation algorithm.

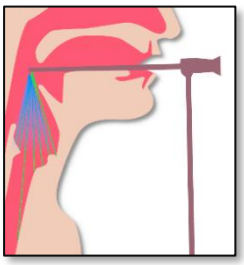


Figure 5 - A symbolic overview of Structured Light Laryngoscopy

Given that we have a calibrated system, a high-speed video recorded with a structured light endoscope as input suffices.

Feature Detection To reliably extract laser-point positions as well as a segmentation of vocal folds and glottal gap, we use the method introduced in [1]. Here, we find laser dots and segmentations jointly by reformulating the task of localizing the laser dots as a segmentation task and finding sub-pixel accurate point positions based on the SoftMax output of a general U-Net trained in a supervised manner.

Correspondence Estimation To estimate correspondences, we first find a single frame in which the glottal gap is minimal, by finding a local minimum of the glottal area waveform.
We assume that the vocal folds are quasi-planar in this state and the amount of detectable laser dots is maximized.
Here, we find initial labeling candidates based on epipolar constraints of the system and domain, e.g. the vocal folds are in general 40mm to 80mm away from the tip of the endoscope.
We then use a RANSAC-based (Random Sample Consensus) recursive hill climbing scheme introduced in [2] to estimate locally optimal correspondences.

Triangulation Based on the previous step, we now can generate a 3D point cloud using general stereo triangulation. Next, we project the glottal outline into the point cloud. At last, we split this point cloud based on the glottal midline (that we can estimate in the Feature Detection step) into left- and right vocal fold.

3D Shape Reconstruction Instead of just fitting a plane into the reconstructed point cloud, we want to reconstruct a complete 3D Mesh. To achieve this, we first model the M5 vocal fold model by Scherer et al. [3] as a B-Spline based model. Next, we use the As-Rigid-As-Possible algorithm [4] to mimic the soft-tissue deformation of vocal folds, i.e. we can use this to also model the movement of the medial surface. As a last step, we compute a least squares fit between the superior surface of the vocal fold model and the reconstructed point cloud on a per frame basis.

Qualitative Results:

Here we show reconstructions of a silicone vocal fold model over time. A quantitative evaluation can be found in [2].

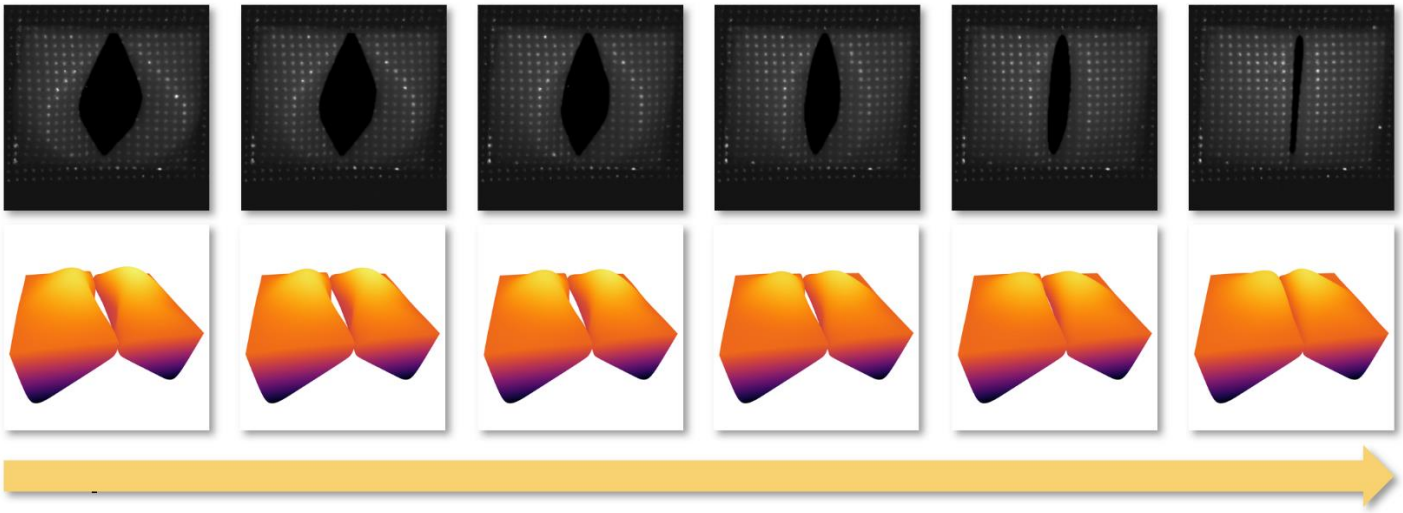


Figure 6 - A 3D reconstruction of an oscillating silicone vocal fold model over multiple frames using our method.

Discussion & Conclusion:

We evaluated our method on a silicone vocal fold model and on in-vivo recordings of the HLE dataset (publicly available here: https://github.com/Henningson/HLEDataset). The reconstruction does not require any human intervention and requires a negligible amount of time (mere seconds).

Since we base our reconstruction on the fitting of a parametric vocal fold model, we can not only visualize the movement of the superior surface, but also visualize the medial axis of oscillating human vocal folds. However, it must be stated that the medial axis is reconstructed in a visually plausible and not physically accurate manner. At the same time, our method can also be used to compute the recording angle, allowing one to give hints about the likelihood of symmetry artifacts stemming from a tilted endoscope. However, we cannot solve ambiguities that stem from the grid-like projection pattern. To reach clinical applicability, other projection pattern need to be considered.

Acknowledgements:

This work was supported by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under grant STA662/6-1, Project-ID 448240908 and (partly) funded by the DFG – SFB 1483 – Project-ID 442419336, EmpkinS. The authors gratefully acknowledge the scientific support and HPC resources provided by the Erlangen National High Performance Computing Center of the Friedrich-Alexander-Universität Erlangen-Nürnberg.

References:

[1] Joint Segmentation and Sub-Pixel Localization in Structured Light Laryngoscopy. MICCAI 2023; Henningson J, Stamminger M, Döllinger M, Semmler M
[2] Real-time 3D Reconstruction of Human Vocal Folds via High Speed Laser Endoscopy. MICCAI 2022; Henningson J, Semmler M, Döllinger M, Stamminger M
[3] Intraglottal pressure profiles for a symmetric and oblique glottis with a divergence angle of 10 degrees. J Acoust Soc Am. 2001 Apr; Scherer RC, Shinwari D, Witt KJD, Zhang C, Kucinschi BR, Afjeh AA
[4] As-Rigid-As-Possible Surface Modelling. Proceedings of EUROGRAPHICS/ACM SIGGRAPH Symposium on Geometry Processing; Sorkinge-Hornung, Olga and Alexa, Marc

FSAI-01: A benchmark dataset for aeroacoustic simulations of human phonation

Stefan Schoder[1], Sebastian Falk[2], A. Wurzinger[1], Alexander Lodermeyer[3], Stefan Becker[4], Michael Döllinger[2], Stefan Kniesburges[2]

[1] Institut für Grundlagen und Theorie der Elektrotechnik, TU Graz, Graz, Austria
[2] Department of Otorhinolaryngology, Head and Neck Surgery, Division of Phoniatrics and Pediatric Audiology, University Hospital Erlangen, FAU Erlangen-Nürnberg, Erlangen, Germany
[3] Department of Process Machinery and Systems Engineering, FAU Erlangen-Nürnberg, Erlangen, Germany
[4] Institute of Fluid Mechanics, Friedrich-Alexander-University Erlangen-Nürnberg, Erlangen, Germany

***Keywords*: e.g. Voice; Fluid-Structure-Interaction; Acoustics; Dataset**

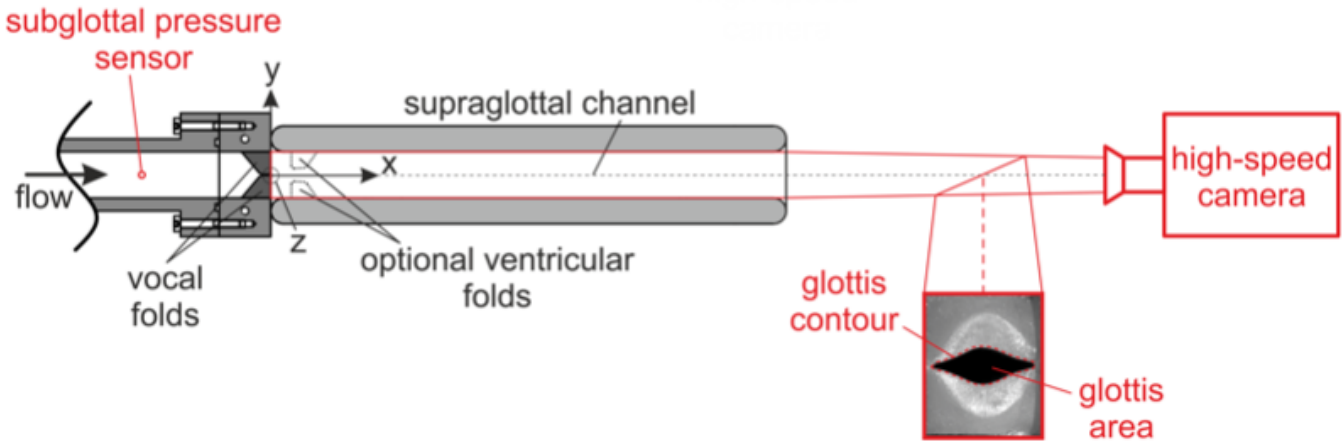
Abstract:
A freely available aerodynamics and aeroacoustic experimental dataset (https://doi.org/10.5281/zenodo.10141715) of the synthetic human phonation model synthVOICE (see Figure 1) is presented [1]. The dataset can be used for benchmarking numerical methods in aerodynamics, aeroacoustics, and the highly complex fluid-structure-acoustic interaction process of voice production.

Objectives / Introduction:
The sound of the human voice is generated by the fluid-structure-acoustic interaction between the laryngeal airflow and the two vocal folds (VFs). To analyze the essential aero-acoustic sound generation, experimental synthetic larynx models were developed to provide reliable and reproducible conditions to investigate the phonatory process accurately. synthVOICE is an internationally well-respected experimental larynx model that includes two VFs made of silicone showing periodical oscillations similar in amplitude and frequency to those found in male phonation. The objective is to present the wide range of experimental data in the online database. This includes the model geometry, the material characteristics, quantities of the vocal fold dynamics, fluid mechanical quantities, and the acoustic field.

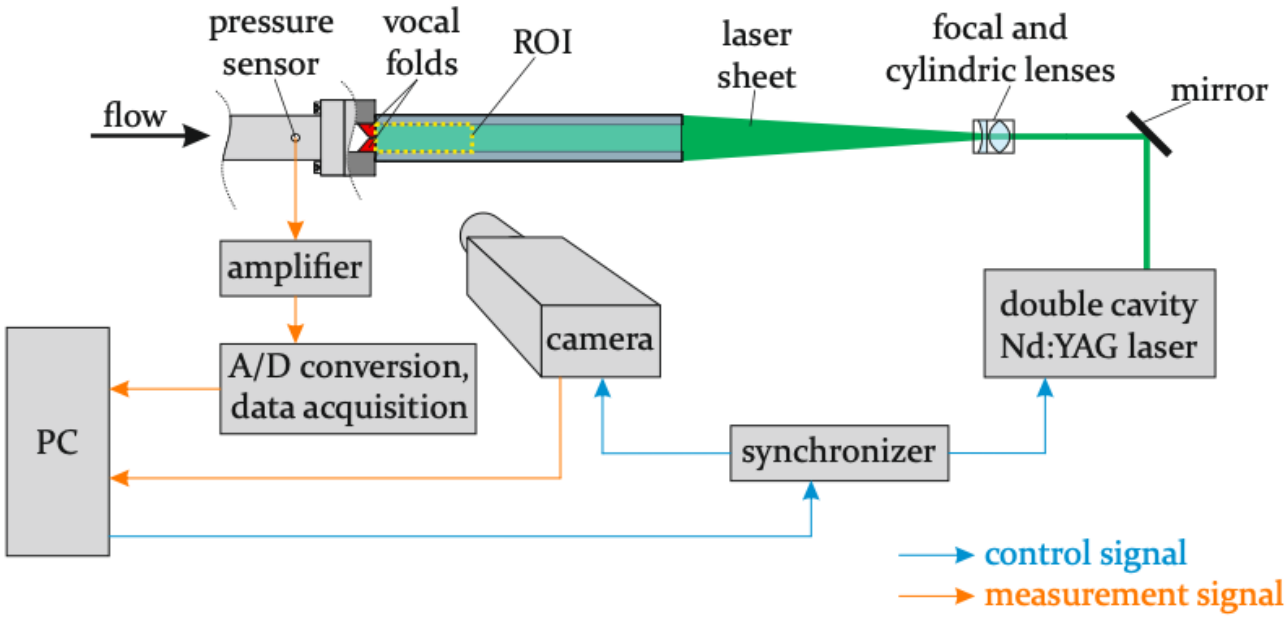
Methods:
The benchmark case was transferred from a synthetic larynx and includes a large amount of experimental data of the fluid flow, the periodic dynamics of the silicone vocal folds and the produced sound field. The experimental setup consists of a mass flow generator, an aeroacoustic silencer, and the synthetic larynx model synthVOICE. synthVOICE consists of two elastic M5 silicone vocal fold models fixed within a mounting frame. The subglottal channel and the mounting frame consist of aluminum. The supraglottal test channel downstream of the vocal folds comprises two transparent walls made of acrylic glass (particle image velocimetry access) and two lateral walls made of aluminum. The two ventricular folds were inserted optionally at a distance of 5mm downstream of the vocal folds. The obtained data from this model are shown in the following table [2].

Data set	VeF	f_0	$\overline{P}_{sub}$	Q	Measuring data
Acoustic &		143 Hz	3251 Pa	100 l/min	sub- & supraglottal pressure @ 40 positions
aerodynamic pressure	✓	148 Hz	2449 Pa	70 l/min	sound pressure far field @ 4 positions
Visualization		145 Hz	3667 Pa	100 l/min	subglottal pressure
VFs dynamics	✓	148 Hz	2754 Pa	75 l/min	high-speed footage
Flow field PIV		146 Hz	3322 Pa	95 l/min	2D flow field mid-coronal plane subglottal pressure

Results:
A large amount of experimental data is available online, including the aerodynamic pressure at multiple positions along the flow region, the acoustic pressure, the time-resolved 2D flow field measured by PIV, and high-speed recordings of the vocal fold oscillations (see previous table). Furthermore, the CAD data of synthVOICE and the exact positions of the measuring hardware ensure the reproducibility of the model. In the following, the setup of the high-speed camera measurements is depicted, including the images of the vocal fold motion. For instance, the glottal area waveform of one oscillation cycle can be derived from these results and used as an input for a phonation simulation, as carried out in [3].



The PIV measurement setup inside the supraglottal channel is depicted as a second data-acquiring example.



The dataset provides the opportunity to validate simulation data and workflows. Exemplarily, the validation of an aeroacoustic human phonation model using the hybrid aero-acoustic simulation model called simVOICE [3] is presented. It combines the flow and acoustic propagation simulations using the CAA solver openCFS [4].

Conclusions:

In conclusion, this benchmarking dataset allows the validation of structural dynamics, aerodynamics, and aeroacoustics of a highly complex fluid-structure-acoustic interaction encountered during phonation like motion. The extensive data packages, including the geometry, the material characteristics, quantities of the vocal fold dynamics, fluid mechanical quantities, and the acoustic field, are available online: https://doi.org/10.5281/zenodo.10141715 .

References:

[1] Schoder, Stefan, et al. " A benchmark case for aeroacoustic simulations involving fluid-structure-acoustic interaction transferred from the process of human phonation." *Acta Acustica* X.Y (2024): Z (in review).

[2] Kniesburges, S., & Schoder, S. (2023). FSAI-01: A benchmark case for aeroacoustic simulations involving fluid-structure-acoustic interaction transferred from the process of human phonation. Zenodo. https://doi.org/10.5281/zenodo.10402984

[3] Falk, Sebastian, et al. "3D-FV-FE aeroacoustic larynx model for investigation of functional based voice disorders." *Frontiers in Physiology* 12 (2021): 616985.

[4] Schoder, Stefan, and Klaus Roppert. "openCFS: Open Source Finite Element Software for Coupled Field Simulation--Part Acoustics." *arXiv preprint arXiv:2207.04443* (2022).

EXPERIMENTAL MODELLING OF THE INFLUENCE OF VOCAL FOLDS COMPLIANCE ON HUMAN VOCAL TRACT ACOUSTIC PROPERTIES

Vojtěch Radolf[1], Jaromír Horáček[1], Anne-Maria Laukkanen[2]

[1] Institute of Thermomechanics, Czech Academy of Sciences, Prague, Czzech Republic

[2] Faculty of Social Sciences, Tampere University, Tampere, Finlland

***Keywords*: Vocal tract acoustics, Artificial vocal folds, Yielding waalls effect**

Abstract:

Objectives / Introduction:

Transfer functions of a vocal tract (VT) are often used as an indicator of acoustic resoonance properties of this filter changing the primary sound source to the voice sound at the lips. These properties are, however, strongly dependent on the boundary conditions of the VT. The input boundary conditions, i.e., the vocal foldss, change the state from open to closed during phonation. Moreover, closure is done by soft tissue instead of a hard walll that is usually considered in simplified simulations. Computational as well as experimental modelling of these two pproblems is not still properly resolved. Therefore, we decided to measure and compute the transfer functions of the VVT model connected to vocal folds replica made of silicone, which is used for experimental investigation of voice in ourr laboratory for several years, see e.g. [1]. For the measurements we used the recently published methodology develooped by Fleisher et al [2]. For computational analysis we used previously published method [3].

Methods:

First, the simplified vocal tract model with circular cross-sections was 3D printed using aan acoustically hard material. The base of its main geometric configuration was taken from a 3D volume model obtainned from CT examination for the Czech spoken vowel [a:], see [4]. However, the VT shape was modified by inccreasing the mouth opening, epilarynx tube prolongation and narrowing of it and increasing the low pharynx volume. TThen, the tuning procedure [5] was applied to reach the resonance frequencies of the VT model as for formants estimmated from the recordings of sung vowel [a:] of the female subject. Thus, the VT model of “opera singing [a:]” was obtaained, see Fig. 1.

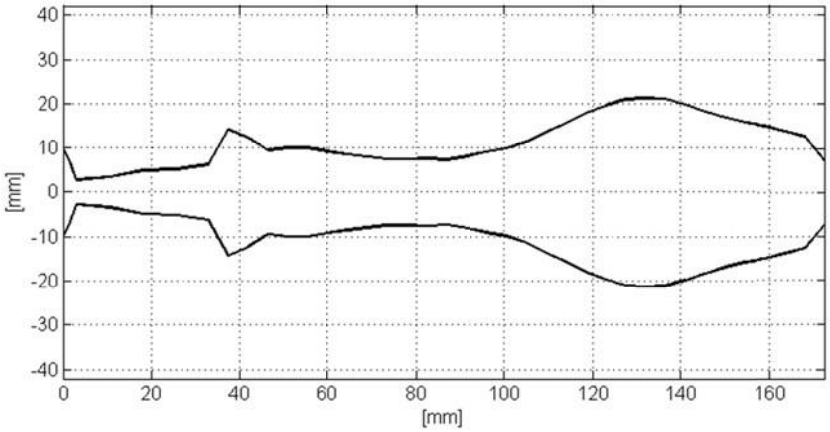


Fig.1: Geometry of the vocal tract model.

The volume velocity transfer function of the VT model was measured using the preccise method described in [2]. This utilizes excitation of the VT model with an external sound source in front of the lips and measurement of the sound pressure inside the models at the glottal end. Thus, the pressure P_1 is measured at the glottis while the mouth is open and then the pressure P_3 is measured right in front of the closed lips. The presssure ratio P_1/P_3 is exactly the volume velocity transfer function, which reflects resonances of the VT open at the lips and closed with a hard wall at the glottis.

However, the input boundary condition of a real human VT is not a hard wall. The vocaal folds created by soft tissue can be very compliant. For modelling we applied the experimental method [2] using the silicone model of vocal folds created from Ecoflex 00-10. The pressure P_1 was measured with a special B&K 41882 microphone probe with a diameter of 1.25 mm. This probe was inserted between the left and right parts of the threee-layer vocal fold (VF) model, which was connected to the VT. The transfer function was then measured for several connditions of VF filled either with pressurized air or water. Changing the medium and its pressure in the middle layer betweeen models of epithelium and body layers inside the VF replica simulated its different stiffnesses.

Results:

Fig. 2 shows the final version of the measured transfer functions, resulting from the aboove-mentioned measurement procedure, for several input boundary conditions (VF stiffness): 1) hard wall instead off VF, 2) VF filled with water $p_{internal}$ = 10 mbar, 3) $p_{internal}$ = 5 mbar, 4) $p_{internal}$ = 0 mbar, 5) VF filled with air $p_{internal}$ = 10 mbar, 6) VF filled with air $p_{internal}$ = 5 mbar, 7) VF evacuated and open. It can be seen that varying the VF stiffneess changes the resonances both in frequency and level. A substantial difference between hard wall and compliant VVF is evident mainly above 2 kHz. The transfer function of the evacuated VF with glottis partially open reveals extra peeaks and dips caused by the model of the subglottal space, see e.g. [1].

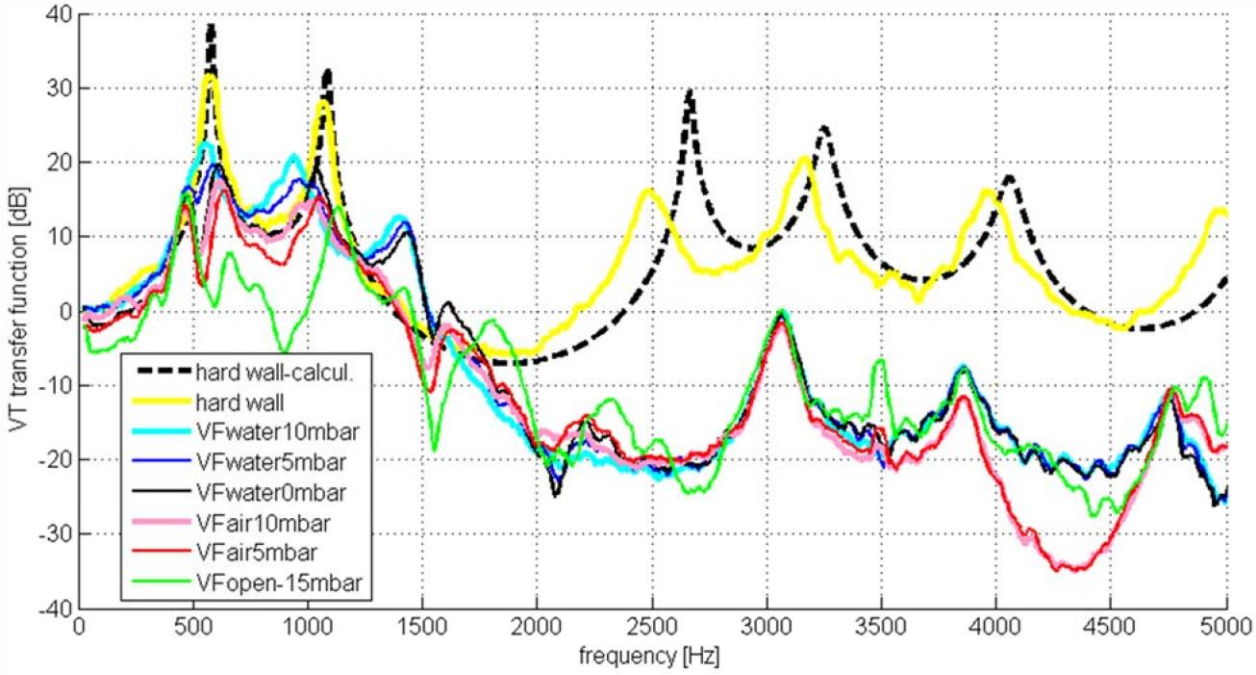


Fig.2: The volume velocity transfer function calculated for hard wall closure at the glottiis and measured for different vocal folds stiffness.

Conclusions:

The influence of VF compliance on the vocal tract transfer function was found to be very important, especially in higher frequency region.

Acknowledgements:

The study was supported by a grant from the Czech Science Foundation: No. 24-111121S and by the Academy of Finland (grant no. 356528).

References:

[1] Laukkanen A-M., Horáček J., Radolf V.: Buzzer versus water resistance phonnation used in voice therapy. Results obtained with physical modeling. BSPC 66, 2021, 102417. https://doi.org/10.10166/j.bspc.2021.102417

[2] Fleischer M., Mainka A., Kürbis S., Birkholz P.: How to precisely measure the voluume velocity transfer function of physical vocal tract models by external excitation. PLoS OONE 13(3): e0193708. https://doi.org/10.1371/journal.pone.0193708

[3] Radolf, V., Horáček, J., Dlask, P., Otčenášek, Z., Geneid, A., Laukkanenn, A.M.: Measurement and mathematical simulation of acoustic characteristics of an artificially lengthened vocal tract. Journal of Sound and Vibration, 366, 2016, 556-570. https://doi.org/10.1016/j.jsv.2015.12.018

[4] Vampola T., Laukkanen A-M., Horáček J., Švec J.G.: Vocal tract changes causedd by phonation into a tube: A case study using computer tomography and finite-element modeling. J. Acoust. Soc. AAm. 129 (1), 2011, 310-315. DOI: 10.1121/1.3506347

[5] Leino T., Laukkanen A-M., Radolf V.: Formation of the Actor's/Speaker's Formantt: A Study Applying Spectrum Analysis and Computer Modeling. J. Voice. 25 (2), 2011, 150-158. doi:10.1016/j.jvoice.20009.10.002

VOCAL FOLD RECONSTRUCTION FROM OPTICAL VELOCITY AND DISPLACEMENT MEASUREMENTS

Christoph Näger[1], Daniel Zieger[2], Tobias Günther[2], Stefan Becker[1]

[1] Institute of Fluid Mechanics, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Bavaria, Germany

[2] Department of Computer Science, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Bavaria, Germany

***Keywords*: e.g. Voice; Synthetic Larynx Model; Source-Filter Interaction; Optical measurements**

Abstract:

Introduction:

The human voice is generated in a complex interplay of fluid flow, structural vibration, and acoustics. In this process, the vocal folds (VF) are stimulated to vibrate by a flow of air from the lungs. This oscillation in turn leads to a modulation of the air flow, resulting in a pulsating jet flow in the vocal tract (VT). The sound that constitutes the voice thereby arises aeroacoustically from the turbulent free jet, as well as vibroacoustically by sound radiation from the vocal fold surface. For a long time, a linear behavior between sound source and filter was assumed, i.e., changes in the source induced by changes in the filter were neglected [1]. However, this simplified representation is not always valid, especially when a resonant frequency of the VT is close to the oscillation frequency of the vocal folds f_o [2]. We want to study the threedimensional vocal folds' oscillatory behavior in a synthetic larynx model with varying vocal tract lengths to study the interaction of vocal fold oscillation and acoustic properties of the vocal tract.

Methods:

The three-dimensional reconstruction of vocal folds in medicine usually involves endoscopy and an approach to extract depth information like structured light or stereo matching of images. The resulting mesh can accurately represent the superior area of the vocal folds, while new approaches also try to reconstruct the inferior area. We want to study the VFs' oscillatory behavior in a synthetic larynx model with varying vocal tract lengths to study the interaction of VF oscillation and acoustic properties of the VT.

Typical methods for a three-dimensional measurement of vocal fold kinematics involve stereo-imaging techniques requiring two elements (either two camera-views or a camera-view and a structured light source). To study the interaction between the vocal tract resonance frequencies and the vocal fold oscillation, our test rig is designed to be able to elongate the VT however. Therefore, with increasing vocal tract length, the angle between the stereo-elements has to become smaller to be able to still access the vocal fold surface, which makes the methods fail eventually. To circumvent this issue, we use a different approach for the three-dimensional vocal fold surface oscillation measurement,

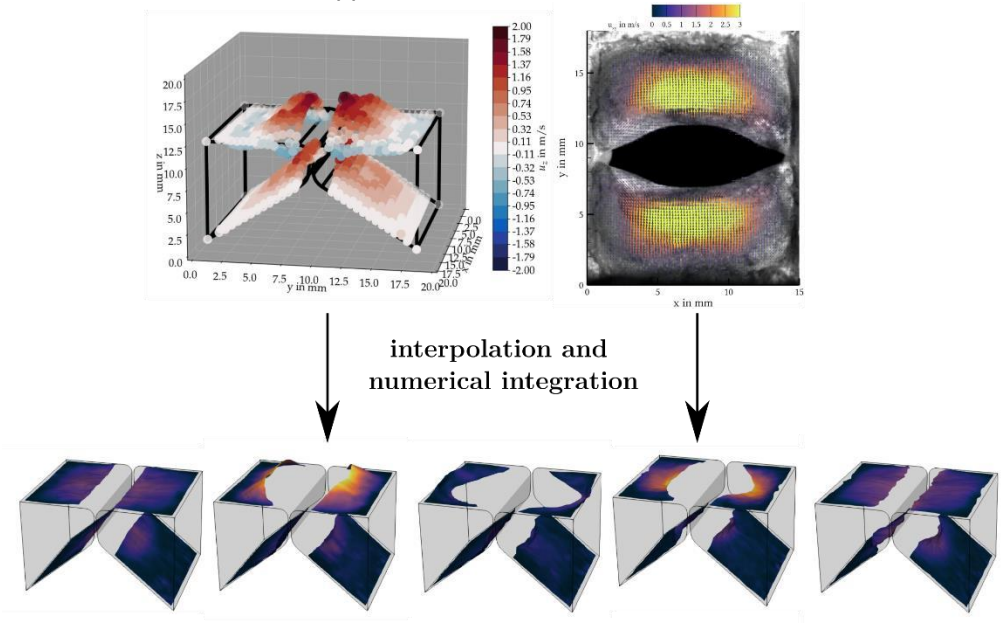


Figure 1 – An overview over the method. Vibrometer and optical flow measurements are interpolated onto a common grid and integrated in time to obtain a three-dimensional displacement field.

combining two different measurement techniques. To measure vocal fold velocities in the transverse plane, a highspeed camera (hsc) recording is made and an optical flow algorithm is applied. Perpendicular to this plane, vocal fold oscillation velocities are measured in the vertical axis via laser-doppler-vibrometry (ldv). For this, a grid containing 702 measurement points is created on the superior vocal fold surface. Both the hsc-recordings as well as the ldvmeasurements are performed with a measurement frequency of 50 kHz. They are synchronized via a subglottal static pressure signal from a flush-mounted pressure sensor in the subglottal channel. As the two methods operate on different measurement grids, a spatial interpolation using radial basis functions is made to obtain surface points with threedimensional velocity vectors. A special curved subglottal channel that allows for optical access of the inferior vocal fold surface is used to be able to apply the procedure to the superior and inferior vocal fold surfaces. With these velocity vectors, a time integration with Romberg's method is made to obtain displacement vectors. These are analyzed in detail with a focus on their harmonic content.

Results:

The method allows us to study the harmonic modes of the vocal fold oscillation on both the inferior and superior vocal fold surface. Five snapshots of one oscillation cycle are shown in Figure 1. Band-pass filtering to single harmonic peaks in the frequency spectrum results in single harmonic oscillation modes. It can clearly be seen that the opening and closing of the vocal folds is a result of the basic oscillation mode at f_o. This mode is also the most relevant for the convergent-divergent glottis movement that is essential for a self-sustained oscillation. Higher harmonics of f_o are dominated by displacements in the vertical axis (z-axis in Figure 1). Changing the vocal tract length can lead to changes in the vocal fold oscillation mode, thereby confirming the presence of acoustic back-coupling onto the vocal fold oscillation [3].

Conclusions:

The method enables us to study the vocal fold oscillation in three dimensions without the need of a stereoscopic arrangement. While the measurements already show promising results, there are still some drawbacks that need to be improved. One problem is the introduction of a slow drift of the displacement in z-axis, resulting from noisy measurement data from the vibrometer. This currently has to be countered by high-pass filtering the data in the z-axis. Another issue is the requirement for a suitable initial positioning of the vocal fold surfaces, from which to integrate the displacement with the velocity measurements. One possible solution could be the introduction of physical constraints to the model in the form of a finite-element-model [4].

Acknowledgements:

This work is funded by the German Research Foundation (DFG) through the project
"Tracing the mechanisms that generate tonal content in voiced speech". Project number: 446965891.

References:

[1] G. Fant: Acoustic Theory of Speech Production. DE GRUYTER, 12 1971

[2] I. R. Titze: Nonlinear source-filter coupling in phonation: Theory. The Journal of the Acoustical Society of America, vol. 123, pp. 2733-2749, 5 2008

[3] F. Kraxberger, C. Näger, M. Laudato, E. Sundström, S. Becker, M. Mihaescu, S. Kniesburges, S. Schoder: On the Alignment of Acoustic and Coupled Mechanic-Acoustic Eigenmodes in Phonation by Supraglottal Duct Variations.
Bioengineering, Vol. 10, No. 12, 2023

[4] D. Zieger, C. Näger, S. Becker, T. Günther: Vocal Fold Reconstruction from Optical Velocity and Displacement Measurements. ArXiv, 2023

EXPERIMENTAL MEASUREMENT OF THE INTERNAL TISSUE VELOCITY IN A SELF-OSCILLATING SYNTHETIC VOCAL FOLD MODEL

Mohsen Motie-Shirazi[1,2], Matías Zañartu[3], Sean D. Peterson[4], Byron D. Erath[1]

[1] Department of Mechanical and Aerospace Engineering, Clarkson University, Potsdam, NY USA
[2] Department of Biomedical Informatics, Emory University, Atlanta, GA USA
[3] Department of Electronic Engineering, Universidad Técnica Federico Santa María, Valparaíso Chile
[4] Department of Mechanical and Mechatronics Engineering, University of Waterloo, Waterloo, ON Canada

***Keywords*: Vocal fold modeling, Internal deformation, Dissipated power**

Abstract:

Objectives / Introduction:

During self-sustained vocal fold (VF) oscillation stresses are imparted to the internal vocal fold tissue. These arise from both tissue deformation due to mucosal wave propagation as well as collision forces when the VFs contact during closing. It is believed that excessive amounts of stress experienced by the tissue, whether due to acute or chronic abuse, is the primary precursor that can lead to the development of structural growths, such as VF nodules.

Despite this understanding, knowledge of the mechanical stresses that arise within VF tissue, particularly during collision, is limited. Although finite element approaches provide obvious advantages for resolving internal tissue motion, validation of these approaches remains a significant concern. *In vivo* measurements have successfully employed ultrasound [1] and optical coherence tomography [2] measurements, although the tissue penetration depth was limited, and both the temporal and spatial resolution were limiting factors. Experimental approaches employing silicone vocal fold models have been successful at resolving surface stresses and strains [3], although measurements of the internal tissue deformation have been more challenging. To the best of our knowledge, only one prior study has pursued this effort, where using magnetic resonance imaging, the displacement of discrete markers embedded in a scaled-up physical vocal fold model was reported [4]. However, the spatial resolution was limited, and the scaling precluded accurate replication of the loading experienced by the *in vivo* vocal folds.

In response, we propose a novel approach that utilizes optically clear synthetic, self-oscillating vocal fold models with embedded fluorescent particles that are illuminated to track internal VF deformations. The utility of this approach is highlighted in this abstract, and measurements relevant to tissue deformation during vocal fold collision are presented.

Methods:

Life-size silicone VF models were fabricated using a layered Body/Cover/Epithelium structure, as shown in Figure 1A. The models were fabricated using optically clear Solaris (Smooth-On) at a part A:part B:thinner ratio of 1:1:4 for the body, 1:1:13 for the cover, and Ecoflex 00-45 Near Clear at a 1:1:0.3 ratio for the epithelium. The resulting moduli of elasticity of the layers are also shown in Figure 1A. Figure 1B shows the optical nature of the Solaris silicone relative to standard Ecoflex 00-30 silicone. When casting each layer, 10.0 μm diameter Fluoro-Max dry red fluorescent particles, which have a peak excitation wavelength of 540 nm and a peak emission wavelength of 612 nm, were mixed into the silicone creating a homogeneous distribution of embedded particles.

The models were affixed to an optically-clear acrylic bracket and placed in a pressure-driven laryngeal flow facility in a hemilaryngeal configuration [5], where they underwent self-sustained oscillations. The subglottal pressure was p_{sg} = 2.40 kPa and the fundamental frequency was f_o = 151 Hz. To measure the internal deformations of the fluorescent particles a LaVision FlowMaster PIV system was used. A Litron Nano L Nd:YAG laser (532 nm, 50 mJ/pulse) was passed through optics to create a 1 mm thick planar laser sheet that was coincident with the mid-coronal VF plane. A sCMOS camera (2,560 x 2,160 pixels) was placed lateral to the VF models to allow visualization of the mid-coronal VF plane. A 550 nm long pass filter was mounted on the end of a Nikon 50 mm f/2.8 lens so that only the illuminated particles inside the VFs were visible, while spurious reflections at the model surface were attenuated. The viewing area and the interrogation scheme (32 x 32 pixel with 50% overlap) resulted in a spatial resolution of 0.1 mm. Image-pairs of the internal deformations were acquired at 20 equally-spaced instances in time throughout one oscillatory cycle. At each instance in time, 50 frame-straddle images were averaged to ensure convergence of the mean displacements.

Results:

Figure 2 shows a sequence of images of the internal material deformation rate (velocity) at three instances throughout the phonatory cycle (t/T = 0.05 (opening), 0.35 (maximum opening), and 0.60 (collision), where T represents the period of oscillation. The complete cycle is not shown for

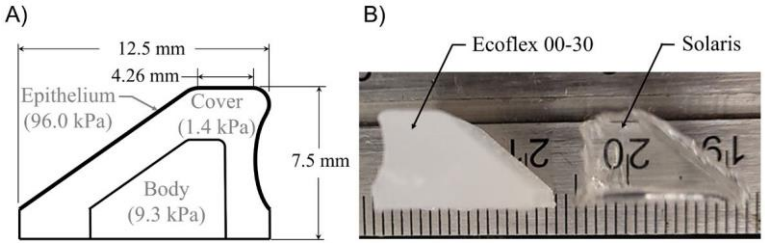


Figure 1: A) Schematic of the VF model with elastic modulus of each layer. B) Traditional versus optically-clear VF models.

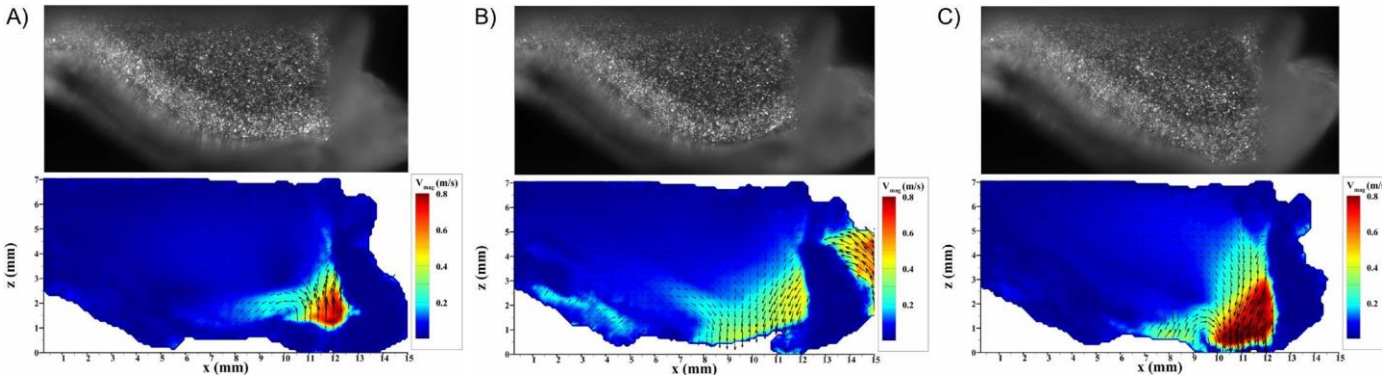


Figure 2: Raw images of the fluorescent particles embedded in the VF models (top row) and the resultant velocity fields acquired using particle image velocimetry (PIV) (bottom row) at times A) t/T = 0.05 (opening), B) t/T = 0.35 (maximum opening), and C) t/T = 0.70 (collision).

brevity. Raw images showing the speckle pattern of the fluorescent particles are shown on top, with interrogated velocity fields below with coloring by velocity magnitude, and velocity vectors overlaid. Note, the difference between the body and cover layers is easily evident in the raw images.

VF motion was primarily confined to the cover layer with the tissue velocity rapidly decreasing in the lateral direction. The majority of the cover motion was in the superior region of the VF motion. During opening (Figure 2A) VF motion was initiated at the superior edge and was highest at the initial opening. Upon reaching maximum opening (Figure 2B) the velocity decreased, and vocal fold closure was initiated at the inferior edge (note the vectors oriented towards the medial direction in Figure 2B). During closure, the tissue velocity was highest immediately preceding closure (Figure 2C). At this moment, the maximum tissue velocity was ~0.8 m/s.

Note, resolution of the internal velocity field was poor along the VF surfaces, as seen in Figure 2. This arose due to both medial and superior VF surface curvature during oscillation. The curved surfaces caused refraction of the imaged particles which is manifest by blurriness in the image. Furthermore, casting the epithelium surface often produced slight wrinkles along the VF surface, which exacerbated this issue.

The kinetic energy per unit volume of the VF was computed along a vertical line that passed from the medial to lateral surface at a position of x = 11 mm, as referenced in Figure 2, at the time of VF collision (t/T = 0.70). Figure 3 shows the velocity magnitude along this line as well as the kinetic energy. Note, both values go to zero at the medial surface (z = 0 mm) due to the loss of information at the surface. As can be seen, the velocity, and more predominantly the kinetic energy, is primarily confined within the cover layer. This information provides useful insights regarding the computation of dissipated power during VF collision, and the subsequent risk of developing phonotraumatic lesions.

Conclusions:

This work has shown the utility of utilizing digital image correlation of embedded fluorescent particles to resolve internal VF kinematics during oscillation and collision, and has shown how it can be used to inform collision dynamics and the estimation of dissipated power. This technique shows significant promise for advancing understanding of internal VF tissue dynamics in both normal and pathological conditions. Furthermore, it provides quantifiable data for validation of numerical models for exploring the etiology of structural VF pathologies.

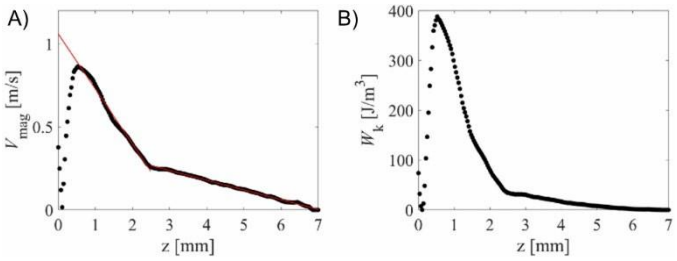


Figure 3: A) Internal velocity, and B) kinetic energy along a vertical line extending in the medial-lateral direction at x = 11 mm.

Acknowledgements:

This research was supported by the National Institutes of Health, National Institute on Deafness and Other Communication Disorders, Grant No. P50 DC015446.

References:

[1] C. G. Tsai, J. H. Chen, Y. W. Shau and T. Y Hsiao, “Dynamic B-mode ultrasound imaging of vocal fold vibration during phonation,” Ultrasound Imaging in Medicine and Biology, vol. 35, pp 1812-1818, 2009.

[2] G. K. Sharma, L. Y. Chen, L. Chou, C. Badger, E. Hong, S. Rangarajan, T. H. Chang, W. B. Armstrong, S. P. Verma, Z. Chen and R. Ramalingam, “Surface kinematic and depth-resolved analysis of human vocal folds in vivo during phonation using optical coherence tomography,” Journal fo Biomedical Optics, vol. 26, pp. 086005, 2021.

[3] M. Spencer, T. Siegmund and L. Mongeau, “Determination of superior surface strains and stresses, and vocal fold contact pressure in a synthetic larynx model using digital image correlation,” The Journal of the Acoustical Society of America, vol. 123, pp. 1089-1103, 2008.

[4] C. J. Taylor, G. J. Tarbox, B. D. Bolster, N. K. Bangerter and S. L. Thomson, “Magnetic resonance imaging-based measurement of internal deformation of vibrating vocal fold models,” The Journal of the Acoustical Society of America, vol. 145, pp. 989-997.

[5] M. Motie-Shirazi, M. Zañartu, S. D. Peterson, B. D. Erath, “Vocal fold dynamics in a synthetic self-oscillating model: Intralgottal aerodynamic pressure and energy,” The Journal of the Acoustical Society of America, vol. 150, pp. 13321345, 2021.

INTRAGLOTTAL PRESSURE AND CONTACT PRESSURE DISTRIBUTION IN EXCISED HUMAN LARYNGES

Sarah Lehoux[1], Zhaoyan Zhang[1]

[1] Dept. of Head & Neck Surgery, University of California, Los Angeles, California, USA

***Keywords*: e.g. Intraglottal pressure; Excised larynges; Contact pressure; Vocal fold thickness**

Abstract:

Objectives / Introduction:

During phonation, the vocal folds (VFs) often vibrate with a vertical phase difference between the upper and lower margins of the medial surface, which alters the shape of the VF medial surface within an oscillation cycle: during the opening phase, the medial shape is convergent, and during the closing phase the medial shape is divergent (or less convergent). This change in the VF medial shape is assumed to facilitate VF self-oscillation by generating a larger pressure during the opening phase than during the closing phase [1-3]. To better understand the mechanisms of VF oscillation, it is important to explore the dynamic distributions of the pressure inside the larynx, and how they change with different laryngeal and supralaryngeal configurations. In particular, the dynamic distribution of the contact pressure along the VF medial surface is relevant in investigating the occurrences of VF injuries and the strategies to prevent them [4-5]. Here we propose to measure the dynamic intraglottal pressure distribution on the medial surface using an excised human larynx. The goal of this preliminary study was to use a small (0.5 mm diameter) probe to obtain the pressure distribution with a fine spatial resolution, and observe its variations with different laryngeal adjustments.

Methods:

One excised human larynx (male, 40 years old) was used for the experiment. The night before the experiment, the larynx was slowly thawed in a fridge at the temperature of -4°C, and on the day of the experiment it was soaked in saline solution. The larynx was mounted on a tube connected to an expansion chamber mimicking the lungs, and airflow was provided through a constant-pressure source. Supralaryngeal superfluous tissues were removed to make the VFs visible, and sutures were attached to mimic the actions of the lateral cricoarytenoid (LCA) and cricothyroid (CT) muscles. Weights were attached to the sutures to simulate different degrees of muscle activation.

The intraglottal pressure was measured by a probe microphone type 4182 (Brüel & Kjær, Nærum, Denmark) with a modified probe similar to the one described by Chen & Mongeau [6]: it was 100 mm long, closed at the end, and a hole was drilled on the side, approximately 20 mm from the closed end. The intraglottal pressure was registered across a grid containing 51 rows and 5 columns, spanning 10 mm vertically (0.2 mm increments from slightly above the highest VF displacement to below the VFs) and 4 mm on the anterior-posterior direction (1 mm increments), for a

duration of 0.25 s each. The mean airflow, mean subglottal pressure, subglottal acoustic pressure, and radiated pressure were also simultaneously recorded. The placement of the probe microphone was controlled and automated by two step motors.

The airflow was adjusted so that the VF oscillation was steady and regular. This allowed for a consistent cycle synchronization based on the subglottal acoustic pressure waveform.

Results:

Figure 1 depicts the pressure distribution recorded by the probe microphone for two contrasting laryngeal configurations: on one hand the aim was to thin and stretch the VFs by simulating a high CT and LCA muscle activation (Fig. 1a), and on the other hand the aim was to make the VFs thick, lax and short by reducing the weights attached to the elongation sutures while pushing the thyroid cartilage backwards (Fig. 1b). As expected in the absence of a vocal tract, the pressure values right above the VFs (top in the figure) is negligible in comparison with the intraglottal and subglottal pressure values. Similarly, as the probe was moving downwards the pressure waveform became more and more similar to the subglottal pressure waveform.

Most interestingly, we observed that in the region between the subglottal and supraglottal space, the phase at which the peak intraglottal pressure occurred varied with the vertical location. This can be seen in Figure 1 as local maxima propagating upwards as time passes. This upward propagating pattern of the peak intraglottal pressure likely resulted from a VF contact region moving upwards with time, which goes in line with a wave-like vibration of the VFs. The extent of the peak travel can therefore be used as an estimation of the effective VF thickness: for the "thin" VFs condition (Fig. 1a), it is roughly 3 mm, whereas in the "thick" condition (Fig. 1b) it is roughly 5 mm.

The pressure distribution did not exhibit substantial variations along the anterior-posterior direction. However, we observed substantial variations of the intraglottal pressure waveform when moving the probe in the medial-lateral direction.

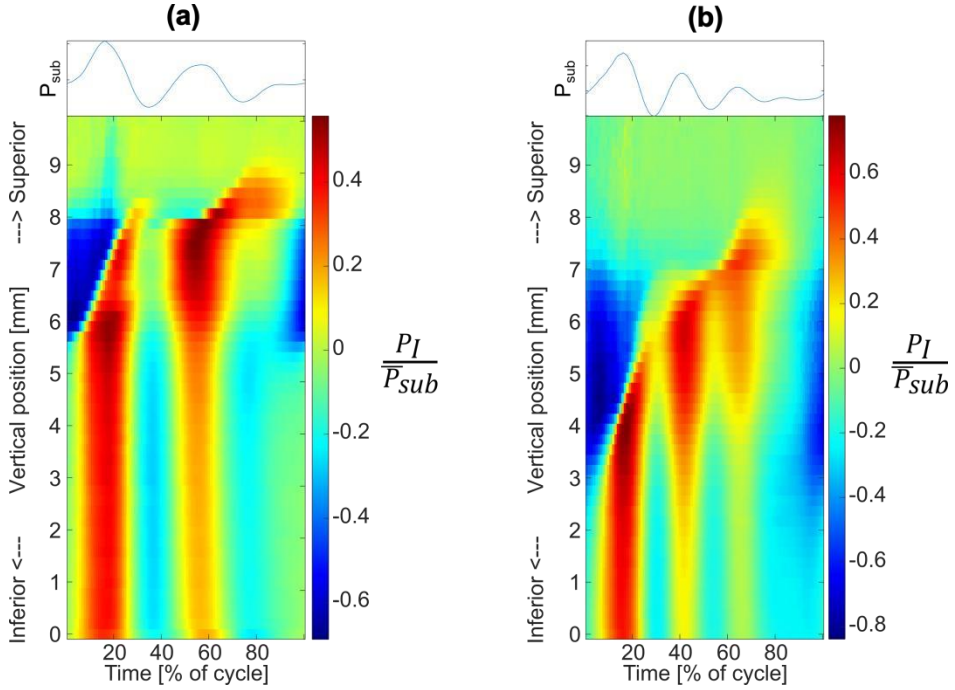


Figure 1. Dynamic distribution of the intraglottal pressure in the vertical dimension within one oscillation cycle. The intraglottal pressure values P_I are normalized by the mean subglottal pressure $\bar{P}_{sub}$. The top panels represent the subglottal acoustic pressure waveform P_{sub}. (a) "thin" VFs condition ($\bar{P}_{sub} = 2.4$ kPa); (b) "thick" VFs condition ($\bar{P}_{sub} = 2.3$ kPa);.

Discussion and conclusions:

This study presents the first fine-resolution, ex vivo measurement of the intraglottal and contact pressure distribution across the medial surface of the VFs, using an excised human larynx. We used a full larynx, where our probe was disrupting the full extent of the VF vibrations, which contrasts with most previous studies that used hemilaryngeal models [4,7]. Nevertheless, we believe that the results of this study provide new insights into the mechanisms of voice production and are relevant for the measurement of the intraglottal and contact pressures in living human subjects, where the conditions are similar to our study.

Acknowledgements:

This work was supported by research grant R01 DC020240 from the National Institute on Deafness and Other Communication Disorders, the National Institutes of Health.

References:

[1] Titze, I. R. (1988). The physics of small-amplitude oscillation of the vocal folds. *J. Acoust. Soc. Am.*, *83*(4), 15361552.

[2] Zhang, Z. (2016). Mechanics of human voice production and control. *J. Acoust. Soc. Am.*, *140*(4), 2614-2635.

[3] Švec, J. G., et al. (2021). Integrative insights into the myoelastic-aerodynamic theory and acoustics of phonation. Scientific tribute to Donald G. Miller. *J. Voice*, *37*(3), 305-313.

[4] Jiang, J. J., & Titze, I. R. (1994). Measurement of vocal fold intraglottal pressure and impact stress. *J. Voice*, *8*(2), 132-144.

[5] Zhang, Z. (2020). Laryngeal strategies to minimize vocal fold contact pressure and their effect on voice production. *J. Acoust. Soc. Am.*, *148*(2), 1039-1050.

[6] Chen, L. J., & Mongeau, L. (2011). Verification of two minimally invasive methods for the estimation of the contact pressure in human vocal folds during phonation. *J. Acoust. Soc. Am.*, *130*(3), 1618-1627.

[7] Mehta, D. D., et al. (2019). Toward development of a vocal fold contact pressure probe: Bench-top validation of a dual-sensor probe using excised human larynx models. *Appl. Sci.*, *9*(20), 4360.

VIBRATORY RESPONSE OF AN ADAPTIVE SYNTHETIC LARYNX MODEL

Bogac Tur[1], Lucia Gühring[1], Olaf Wendler[1], Samuel Schlicht[2], Dietmar Drummer[2], Anne Schützenberger[1], Stefan Kniesburges[1]

[1] Division of Phoniatrics and Pediatric Audiology, Department of Otorhinolaryngology, Head and Neck Surgery, University Hospital Erlangen, Medical School, Friedrich-Alexander-Universität Erlangen-Nürnberg, Waldstrasse 1, 91054 Erlangen, Germany

[2] Institute of Polymer Technology, Friedrich-Alexander-Universität Erlangen-Nürnberg, Am Weichselgarten 10, 91058 Erlangen, Germany

***Keywords*: Synthetic vocal folds; Vibratory response; Larynx dynamic; Biomimetics**

Abstract:

Objectives / Introduction:

Phonation is a complex physical interplay involving the vocal fold tissue, which is set into oscillation by the airflow, and the resulting acoustics. Synthetic vocal folds and larynx models play a crucial role in fully understanding this intricate fluid-structure-acoustic interaction. Additionally, such synthetic models can replicate various voice pathologies and can serve as preliminary work for larynx implants. This study presents a biomimetic synthetic vocal fold and larynx model capable of reproducing various laryngeal dynamics and adaptive stiffnesses of the vocal folds.

Methods:

The vocal folds are composed of silicone and a three-layer vocal fold model with integrated fibers. The study examines the impact of varying fiber pretension levels and increasing flow rates, while having a pre phonatory closed glottis. Furthermore, as the flow increased, the adduction of the vocal folds was incrementally increased with fiber tension ranging from 0 to 10 mm. The measurements utilize a multimodal approach, including subglottal pressure, vocal fold dynamics, and supraglottal acoustic signals, with adduction controlled through manipulators.

Results:

The models response to changes in fiber tension (see Fig. 1) demonstrates that the fundamental frequency increases with increasing fiber tension at a constant airflow. Additionally, we observed an increase in the fundamental frequency with increasing airflow at constant tension.

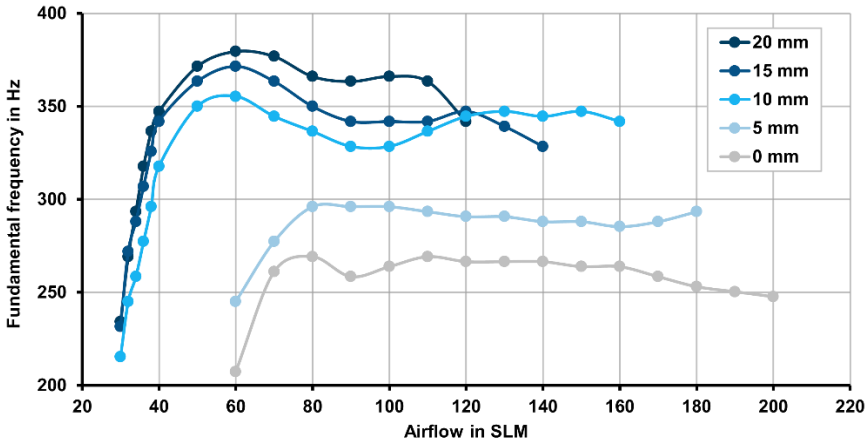


Figure 7: The plot shows the influence of different fiber tensioning levels on the fundamental frequency in Hz as a function of the airflow in SLM

Regarding adduction (see Fig. 2), the model exhibits two different phonation modes: the first at approximately 150 Hz – 200 Hz and the second at 260 Hz – 325 Hz. In both modes, it is evident that adduction decreases with higher airflow to maintain stable phonation in the respective mode.

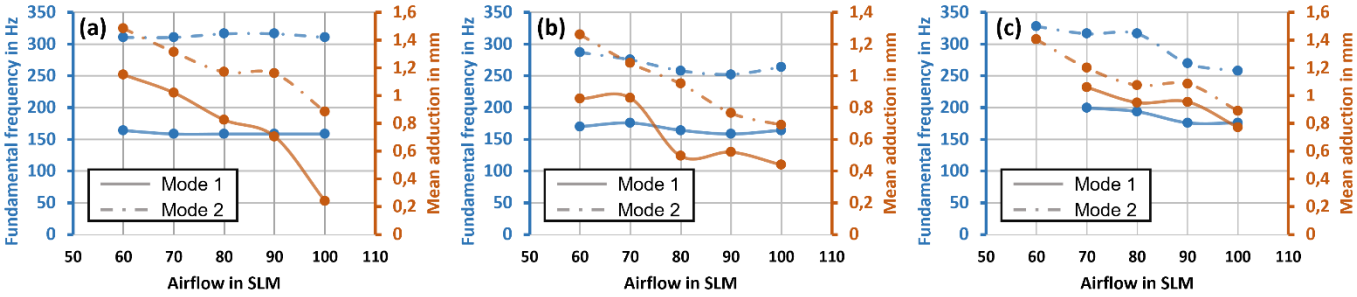


Figure 8: The figure shows the relationship between adduction in mm (orange curves) and the resulting fundamental frequency in Hz (blue curves) as a function of the airflow in SLM for phonation mode 1 (solid line) and phonation mode 2 (dashed line). The vibratory response was tested for fiber tensions at **(a)** 0 mm, **(b)** 5 mm, and **(c)** 10 mm.

Conclusions:

The vibratory response of the model can be adaptively adjusted by manipulating the airflow, fiber tension, and adduction, allowing for significant physiological and pathological variability in the phonation process.

References:

[1] Tur B, Gühring L, Wendler O, Schlicht S, Drummer D, Kniesburges S. Effect of Ligament Fibers on Dynamics of Synthetic, Self-Oscillating Vocal Folds in a Biomimetic Larynx Model. Bioengineering. 2023; 10(10):1130. https://doi.org/10.3390/bioengineering10101130

Highly dynamic rotational molding of thin-walled larynx models for fluid dynamic modelling

Samuel Schlicht[1], Anke Kaufmann[1], Bogac Tur[2], Stefan Kniesburges[2], Dietmar Drummer[1]

[1] Institute of Polymer Technology, Friedrich-Alexander-Universität Erlangen-Nürnberg, Am Weichselgarten 10, 91058 Erlangen

[2] Division of Phoniatrics and Pediatric Audiology at the Department of Otorhinolaryngology Head & Neck Surgery, University Hospital Erlangen, Waldstrasse 1, 91054 Erlangen, Germany

***Keywords*: Laryngeal modeling, rotational molding, fluid-structure coupling**

Abstract:

Objectives / Introduction:

The experimental investigation of the flow-structure coupling represents a significant challenge in otolaryngology/phoniatrics, which is often performed with human and animal-derived specimens showing a rapid degradation of relevant mechanical and dynamic properties. The synthetic, elastomer-based replication of laryngeal mechanics remains challenging due to the complex, non-linear mechanical properties of the larynx. Complementing previously described approaches relying on the multi-layered molding of laryngeal structures, the presented approach is based on the adhesion-controlled highly dynamic rotational molding of ultra-soft silicones, yielding film thicknesses as low as 10 micrometers with ultra-soft properties and high geometric precision.

Methods:

Relying on rotational frequencies exceeding 30 Hz, the manufacturing of thin-walled laryngeal models becomes feasible. Based on the multi-layered manufacturing, the targeted modification of superficial mechanics with ultra-soft properties is obtained while enabling the replication of complex geometries and undercuts. The subsequent elastomer- and hydrogel-based modification of molded silicone structures allows for adapting the mechanical and dynamic properties of the vocal folds, representing a promising approach for the facilitated modeling of laryngeal dynamics.

Results:

Based on the described method, first prototypes of vocal fold models were manufactured which are shown in fig. 1. These prototypes were evaluated in a customized experimental setup by measuring the vocal folds' dynamics, the subglottal pressure and the generated sound signal.

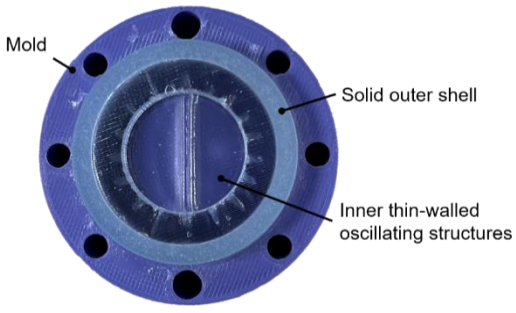


Figure 9: Depiction of rotationally molded laryngeal model inside the additively manufactured mold

Conclusions:

While the manufacturing of ultra-soft laryngeal models with a high geometric precision is usually impaired due to material-specific limitations of ultra-soft polymers, employing the multi-layered manufacturing of elastomers of varying mechanical properties allows to mimic the biomechanical properties of the vocal folds while minimizing adverse adhesive effects and the aging of models.

Acknowledgements:

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - Project-ID 401501386 “Doppelroboteranlage für additive Fertigung”.

References:

N.A.

FLOW-INDUCED OSCILLATIONS OF A VOCAL-FOLD REPLICA WITH TAILORED ANTERIOR-POSTERIOR AND MEDIAL-LATERAL ARTICULATIONS

Mounib Tlaidi[1,2], Sylvain Arnaud[1], Lucie Bailly[2], Nathalie Henrich Bernardoni[1]

[1] Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, 38000 Grenoble, France
[2] Univ. Grenoble Alpes, CNRS, Grenoble INP, 3SR, 38000 Grenoble, France

***Keywords*: Mechatronic larynx testbed; Articulated vocal-fold replica; Glottal vibrations; Aero-acoustic measurements;**

Abstract:

Objectives / Introduction:
In the past two decades, *in vitro* studies of human phonation have shifted from static or controlled oscillations to self-oscillating bio/phono-mimetic replicas, yielding promising results (1–5). Artificial vocal folds have high availability, affordability, and they can be reused during an extended period of time. Moreover, control and access to the vibration and airflow parameters is made easier. Finally, they are essential to better evaluate and link together the results that come from theoretical and numerical models (2). However, they are generally still far from representing the physiological reality of human vocal folds. Besides, few studies have focused on varying either the mechanical properties of the chosen material or the pre-phonatory strain applied to the replica to better understand their influence on its vibratory properties. In particular, only four studies have investigated the impact of changing the vocal-fold anterior-posterior strain condition (ε_{ap}) on fundamental frequency of oscillation (f_0), and on subglottal pressure-airflow relationship (4–7). Owaki *et al.* (7) used a rubber model of vocal folds to measure the impact of material stiffness and elongation. Shaw *et al.* (6) compared silicone-based vocal-fold models with and without embedded fibers, highlighting the impact of material anisotropy on strain-induced variations in f_0 and pressure onset (P_{on}). Tur *et al.* (4) designed a multilayer fiber-reinforced silicone model as a close representation of human vocal folds, exploring the effects of changing diameter, resistance and tension of the ligament fibers, which resulted in a very promising frequency range of 200-520 Hz. Finally, Luizard *et al.* (5) developed an extensible vocal-fold replica with tunable material properties (using silicones or gelatin-based hydrogels), and evidenced a relationship between f_0 and fold stiffness. Yet, the range of f_0 was much lower than expected for human speech, and it did not vary much with longitudinal elongation, calling for further investigation.
In this paper, we present an improved version of the testbed developed in Luizard *et al.* (5), allowing both anterior-posterior elongation and medial-lateral compression. Flow-induced oscillations were characterized on a homogeneous silicone model with varying longitudinal and lateral strain conditions, and airflow rate.

Methods:

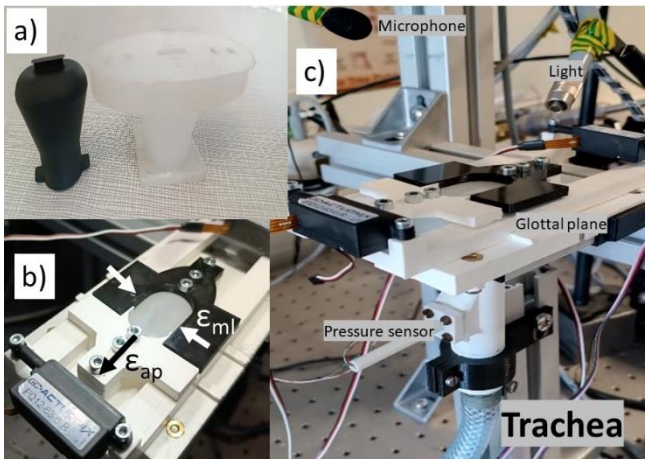


Figure 1: (a) Vocal folds and the 3D-printed inner mold, (b) Applied pre-strains, (c) General view of the testbed

Vocal fold replica. An existing articulated larynx testbed with folds actuation in longitudinal stretching was used as a starting block (5). Improvements were made on the artificial larynx and its actuation in lateral compression. Laryngeal envelope and vocal folds were molded into a single block of homogeneous silicone rubber of shore hardness 00-30. The folds' length at rest was 12mm and their thickness was 1.5mm. Their geometry was inspired from M5 models (8), corresponding to a simplified two-circle version of Microvoice model (5).

Vibroacoustic characterization. The laryngeal replica was loaded by adjustable strains in the anterior-posterior direction (ε_{ap}) and medio-lateral one (ε_{ml}). Airflow rate (ϕ) was controlled by means of an electromechanical valve actuator, as in (5). Several parameters characterizing the oscillatory behavior were measured: subglottal pressure (P_{sg}), fundamental frequency of oscillation (f_o), sound pressure level (SPL).

Results:

As shown in Figure 2a, an increase of 0.30 in longitudinal strain ε_{ap} results in an increase of 40 Hz in fundamental frequency f_o of oscillation within the whole airflow range. This qualitative trend is consistent with (4, 6) in case of fiber-reinforced folds. Previous vocal-fold models with homogeneous and isotropic material properties such as those used in (5, 6) did not show an increase in frequency f_o with the applied anterior-posterior pre-strain.

For a given pre-strain, varying the mean airflow rate has minimal effect on f_o , which varies at most in the range of 10 Hz, consistent with (5).

As evidenced in Figure 2b, the mean subglottal pressure increases non-linearly with the mean airflow rate ϕ up to 2L/s (depending on the pre-strain level), then reaching a stable state. On the other hand, pressure decreases with increasing pre-strain, contrary to (4,5) but in line with (6) who found a decreasing onset pressure with increasing length for linear models.

These discrepancies between previous articulated larynx testbeds are attributed to the critical role played by the boundary conditions applied to the folds, which differ particularly in the medial-lateral and inferior-superior directions.

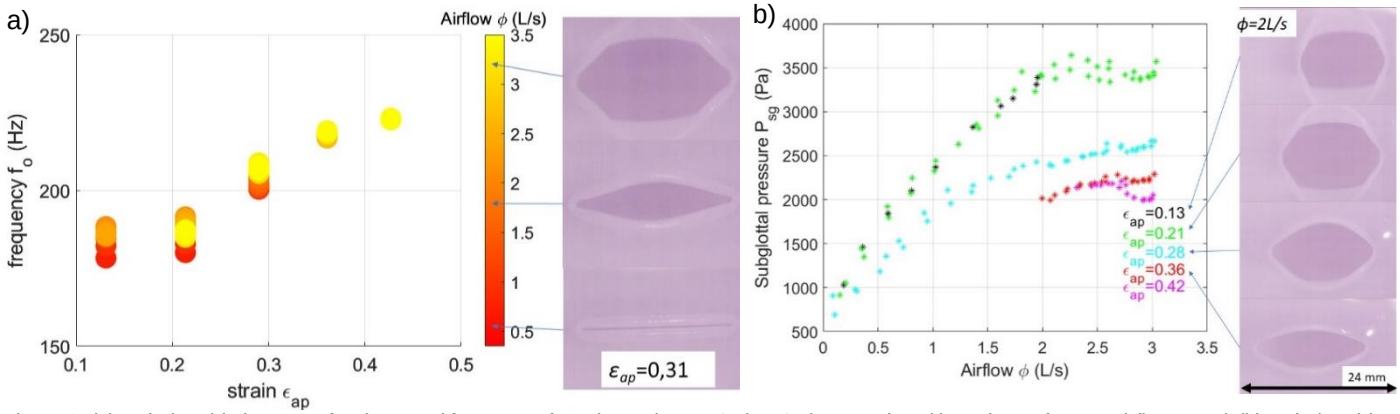


Figure 2: (a) Relationship between fundamental frequency f_o and anterior-posterior strain ε_{ap}, colored by values of mean airflow rate ϕ.(b) Relationship between mean subglottal pressure P_{sg} and ϕ, colored by strain ε_{ap}. Fold's snips correspond to maximum glottal apertures.

Conclusions:

This study introduces an articulated *in vitro* larynx testbed, allowing a quantitative control over both anterior-posterior pre-elongation and medial-lateral pre-compression. These developments show encouraging results on the f_o-variations obtained by activating such articulations, while demonstrating the importance of boundary conditions for accurate replication of previous aero-acoustic results.

Acknowledgements:

This work was supported by the French National Research Agency (ANR AVATARS Grant No. ANR-22-CE48-0014).

References:

1. Kniesburges S, L. Thomson S, Barney A, Triep M, Sidlof P, Horacek J, et al. In Vitro Experimental Investigation of Voice Production. Curr Bioinforma. 1 sept 2011;6(3):305-22.
2. Döllinger M, Zhang Z, Schoder S, Šidlof P, Tur B, Kniesburges S. Overview on state-of-the-art numerical modeling of the phonation process. Acta Acust. 2023;7:25.
3. Häsner P, Birkholz P. Reproducibility and Aging of Different Silicone Vocal Folds Models. J Voice. mars 2023;S0892199723000851.
4. Tur B, Gühring L, Wendler O, Schlicht S, Drummer D, Kniesburges S. Effect of Ligament Fibers on Dynamics of Synthetic, Self-Oscillating Vocal Folds in a Biomimetic Larynx Model. Bioengineering. 26 sept 2023;10(10):1130.
5. Luizard P, Bailly L, Yousefi-Mashouf H, Girault R, Orgéas L, Henrich Bernardoni N. Flow-induced oscillations of vocal-fold replicas with tuned extensibility and material properties. Sci Rep. 19 déc 2023;13(1):22658.
6. Shaw SM, Thomson SL, Dromey C, Smith S. Frequency Response of Synthetic Vocal Fold Models With Linear and Nonlinear Material Properties. J Speech Lang Hear Res. oct 2012;55(5):1395-406.
7. Owaki S, Kataoka H, Shimizu T. Relationship Between Transglottal Pressure and Fundamental Frequency of Phonation—Study Using a Rubber Model. J Voice. mars 2010;24(2):127-32.
8. Murray PR, Thomson SL. Vibratory responses of synthetic, self-oscillating vocal fold models. J Acoust Soc Am. 1 nov 2012;132(5):3428-38.

VOICE ONSET AND OFFSET OSCILLATORY PATTERNS BASED ON PLACE, MANNER AND VOICED/VOICELES CONSONANT ENVIRONMENT

Melda Kunduk[1,2,3], Takeshi Ikuma[1,2], Andrew J McWhorter[1,2], Tobias Schraut[4], Michael Doellinger[4], Robin Samlan[5]

[1]Dept. of Otolaryngology-Head and Neck Surgery, LSU Health Sciences Center, New Orleans, Louisiana, U.S.A.
[2]Voice Center, The Our Lady of The Lake Regional Medical Center, Baton Rouge, Louisiana, U.S.A.
[3]Dept. of Communication Sciences & Disorders, Louisiana State University, Baton Rouge, Louisiana, U.S.A.
[4]Dept. of Otorhinolaryngology Head & Neck Surgery, Medical School Division of Phoniatrics and Pediatric Audiology, University Hospital Erlangen, Erlangen, Germany
[5]Dept. of Speech, Language, & Hearing Sciences, University of Arizona, U.S.A.

Keywords: *Voice; Highspeed Videoendoscopy; Voice onset and offset oscillatory patterns*

Abstract

Objectives / Introduction:

During speech production, the vocal folds (VFs) undergo almost continuous adduction and abduction to generate the voiced and unvoiced segments. Each voiced segment consists of vibratory patterns that include onset and offset and possibly a short duration of vibration. The mechanics and kinematics of the vocal onset and offset of oscillations potentially contain rich information that can be related to vocal health, dysfunction, and patients' specific voice symptoms such as vocal fatigue and handicap (1,2,3). This presentation demonstrates effects of English consonants varied by their place, manner, and voicing on the onset and offset VF oscillation patterns in a vowel-consonant-vowel (VCV) non-word syllable.

Methods:

***Data*:** High speed recordings (HSV) obtained with nasoendoscopy were used to investigate the onset and offset vibratory patterns during the VCV productions, with consonants varying by their place (bilabial, labiodental, alveolar, labiodental, and velar), manner (stop and fricative), and voicing (voiced/voiceless) features (see Table 1). Minimum of six productions of each VCV (/ipi/, /ibi/, /ifi/, /ivi/, /iti/, /idi/, /isi/, /izi/, /iki/, /igi/) were obtained while a female subject produced them with her habitual voice, resulting in 96 HSV recordings total at 4000 frames/second (fps). The rate of the VCV production was regulated by a metronome to ensure uniformity across recordings (0.92 seconds per production). HSV images and acoustic data were recorded simultaneously.

***Procedures*:** HSV recordings were temporally segmented using the acoustic data as a guide, and each VCV production was individually saved to a video file to play back at 20 fps (i.e., slowed down by 200). Each video was perceptually evaluated around its consonant for the presence or absence of three outcome measures: Arytenoid Abduction (separation), Vocal Fold Vibration Cessation, and Vocal Fold Contact Loss.

***Analysis*:** The data were analyzed in 3 two-factor groups while fixing either manner or place: voicing and place effects on stops (Group 1, N:36); voicing and place effects on fricatives (Group 2; N:28); and voicing and manner effects on alveolar consonants (Group 3, N:28). This arrangement avoids a missing-data problem due to English lacking some consonants (i.e., /ɸ/, /β/, /p̪ /, /b̪ /, /x/, and /ɣ/). The numbers of samples in VCV groups were matched to the minimum available. Those with more samples were randomly sampled for each group.

Table 1: Investigated Vowel-Consonant-Vowel Patterns and Analysis Groupings

Manner	Voicing	Place				
		Bilabial	Labiodental	Alveolar	Velar	
Stop	Voiceless	/ipi/		/iti/	/iki/	⇦ *Group 1*
	Voiced	/ibi/		/idi/	/igi/	
Fricative	Voiceless		/ifi/	/isi/		⇦ *Group 2*
	Voiced		/ivi/	/izi/		

⇧ *Group 3*

Results:

Group 1: Voicing and place effects for stops: All voiceless stops (/ipi/, /iti/, /iki/; N:18) (Table 2) presented with arytenoid abduction, vocal fold vibration cessation, and vocal fold contact loss across alveolar, bilabial, and velar places of articulation before the stop consonant in VCV production. None of the voiced stops (/ibi/, /idi/, /igi/; N:18) (Table 2) consonants demonstrated arytenoid abduction before stopping consonant production. Vocal fold vibration cessation was present only with voiced bilabial and alveolar stops (/ibi/, /idi/); both were present only once in six productions (17%). Vocal fold contact loss happened in 39% of all voiced stops while 83% of voiced bilabial and 33% of voiced alveolar stop productions presented this feature (Table 2).

Group 2: Voicing and place effects for fricatives: All voiceless fricative production (/ifi/, /isi/; N:14) (Table 3) presented arytenoid abduction, vocal fold vibration cessation, and vocal fold contact loss for labiodental and alveolar places of articulation before the fricative production in VCV production. None of the voiced fricatives (/ivi/, /izi/; N:14) demonstrated arytenoid abduction before the fricative in VCV production. Vocal fold vibration cessation was present only in voiced alveolar fricatives (/izi/; N:7) indicating place of articulation effect during VCV production. None of the voiced labiodental fricatives demonstrated this gesture in VCV production. Vocal fold contact loss occurred in all alveolar voiced fricatives (/izi/; N:7) and only 14% voiced labiodental fricatives (/ivi/; N:7) indicating a place of articulation effect during VCV production (Table 3).

Group 3: Voicing and manner effects on alveolar place of articulation: While all voiceless alveolar stops and fricatives (/iti/, /isi/; N:14) demonstrated vocal fold contact loss before the consonant production, only 43% of all voiced alveolar stops (/idi/) and 100% voiced alveolar fricatives (/izi/) did so in VCV production (Table 4).

Table 2: Group 1 / *Voicing and Place Effects* on Stop Consonants

Voicing	Place	Arytenoid Abduction	VF Vibration Cessation	Vocal Fold Contact Loss
Voiceless	(all)	18/18 (100%)	18/18 (100%)	18/18 (100%)
Voiced	Bilabial	0/6 (0%)	1/6 (17%)	5/6 (83%)
	Alveolar	0/6 (0%)	1/6 (17%)	2/6 (33%)
	Velar	0/6 (0%)	0/6 (0%)	0/6 (0%)

Table 3: Group 2 / *Voicing and Place Effects* on Fricative Consonants

Voicing	Place	Arytenoid Abduction	VF Vibration Cessation	Vocal Fold Contact Loss
Voiceless	(all)	14/14 (100%)	14/14 (100%)	14/14 (100%)
Voiced	Labiodental	0/7 (0%)	0/7 (0%)	1/7 (14%)
	Alveolar	0/7 (0%)	7/7 (100%)	7/7 (100%)

Table 4: Group 3 / *Voicing and Manner* Effects on Alveolar Consonants

Voicing	Manner	Arytenoid Abduction	VF Vibration Cessation	Vocal Fold Contact Loss
Voiceless	(all)	14/14 (100%)	14/14 (100%)	14/14 (100%)
Voiced	Stop	0/7 (0%)	1/7 (14%)	3/7 (43%)
	Fricative	0/7 (0%)	7/7 (100%)	7/7 (100%)

Conclusions:

Preliminary results strongly suggest main effects of voicing, place, and manner and interaction effects between voicing and place and between voicing and manner on the vocal fold vibratory patterns of the VCV productions. Arytenoid abduction (separation) with cessation of vibration and loss of vocal fold contact always occurs during VCV production with all of the voiceless consonants assessed (/ipi/, /iti/, /iki/, /ifi/, /isi/). As the arytenoids abduct, VFs have to lose contact, and there is too much air leakage to sustain the vibration. None of the voiced consonants (/ibi/, /idi/, /igi/, /ivi/, /izi/) exhibited arytenoid separation. The vocal fold vibration cessation before the voiced consonant production was observed mostly for alveolar fricative (/izi/) in VCV position. The vocal fold contact loss was observed in some of the voiced alveolar stops but present in all voiced alveolar fricatives. Behaviors of the vocal fold (contact/cessation) with the voiced consonants is dictated by the aerodynamic coupling of the source and vocal tract. Interestingly, the stops and fricatives draw the polar opposite effect of the placement. Stops further from source cause more contact loss while fricatives further from source cause less contact loss and cessation. At the alveolus, the voiced fricative is more likely to cause disruption to the vibration than voiced stop. Indications for future studies and clinical implications of the results will be discussed.

References

1. M. Kunduk, T. Ikuma, D. C. Blouin, and A. J. McWhorter, “Effects of volume, pitch, and phonation type on oscillation initiation and termination phases investigated with high-speed videoendoscopy,” J. Voice, vol. 31, no. 3, pp. 313–322, 2017, doi: 10.1016/j.jvoice.2016.08.016.
2. R. F. Orlikoff, D. D. Deliyski, R. J. Baken, and B. C. Watson, “Validation of a glottographic measure of vocal attack,” J. Voice, vol. 23, no. 2, pp. 164–168, 2009, doi: 10.1016/j.jvoice.2007.08.004.
3. C. E. Stepp, “Relative fundamental frequency during vocal onset and offset in older speakers with and without Parkinson’s disease,” J. Acoust. Soc. Am., vol. 133, no. 3, pp. 1637–1643, Mar. 2013, doi: 10.1121/1.477620

EVALUATION OF INTRINSIC LARYNGEAL MUSLCE ACTIVITY USING SURFACE ELECTROMYOGRAPHY

Xinyi Yang, Hanting Zhang, Liang Wu
Department of Biomedical Engineering, School of Life Science and Technology,
Xi'an Jiaotong University, Xi'an, Shaanxi, China

***Keywords*: Speech control; Surface electromyography; Intrinsic laryngeal muscle activity; Frequency-domain feature**

Abstract:

Objectives / Introduction:

Healthy function of intrinsic laryngeal muscles (ILMs) is extremely important for producing intelligible and highquality speech. This work aims to correctly measure ILM activities for accurate assessment of laryngeal function, which is helpful to further understand neuromuscular control of voice production and to diagnose laryngeal disorders in clinical applications.

Methods:

A non-invasive method using two-channel surface electromyography (sEMG) was designed to extract the activity signals of cricothyroid (CT) and lateral-cricoarytenoid (LCA) muscles through specific sensor placement and signal processing. The activities of CT and LCA muscles were quantitatively evaluated by two time-domain features (i.e., root mean square RMS, mean absolute value MAV) and three frequency-domain features (i.e., median frequency MDF, mean frequency MNF, and frequency ratio FR) of the extracted sEMG signals. The sEMG features were further compared with acoustic features (i.e., fundamental frequency F0 and closed quotient CQ) of simultaneous speech signals to investigate dynamic changes in muscle activity during speech control.

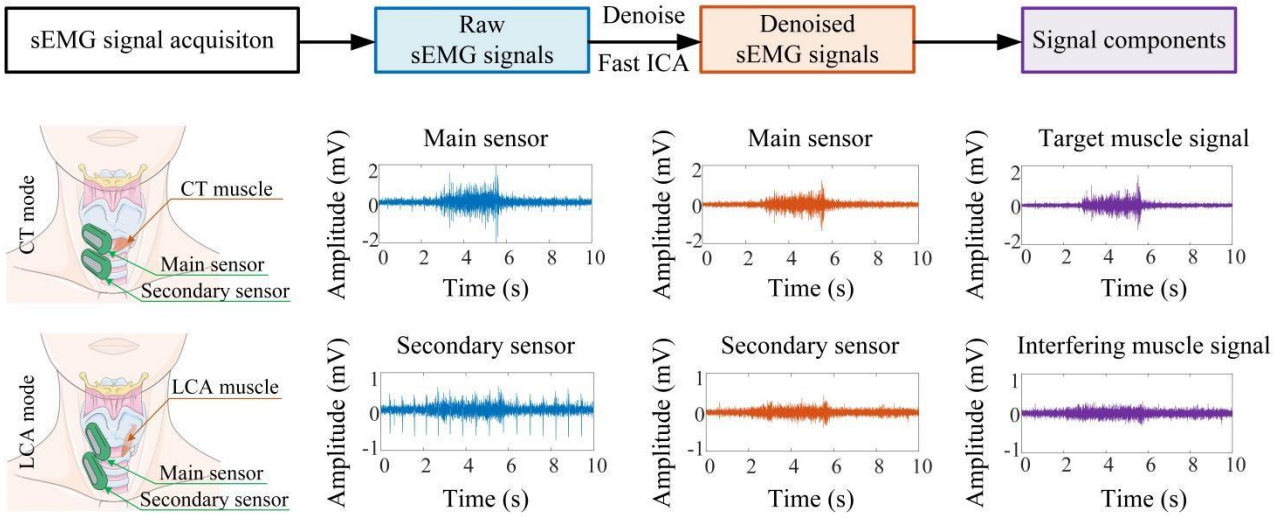


Fig. 1. The sEMG signal collection and processing.

Results:

The frequency-domain features of sEMG signals from CT and LCA muscles had a significant and consistent correlation with F0 but no significant correlation with CQ, indicating an increasing activity of CT and LCA muscles with voice pitch rising. Furthermore, CT muscle showed greater changes in the frequency-domain sEMG features during pitch control than LCA muscle. This outcome was consistent with the effect of CT and LCA muscles on voice control reported in previous experimental studies. Moreover, the frequency-domain features of sEMG signals had a more stable pattern than the time-domain features across subjects, showing potential as reliable parameters for quantitative evaluation of laryngeal muscle activity.

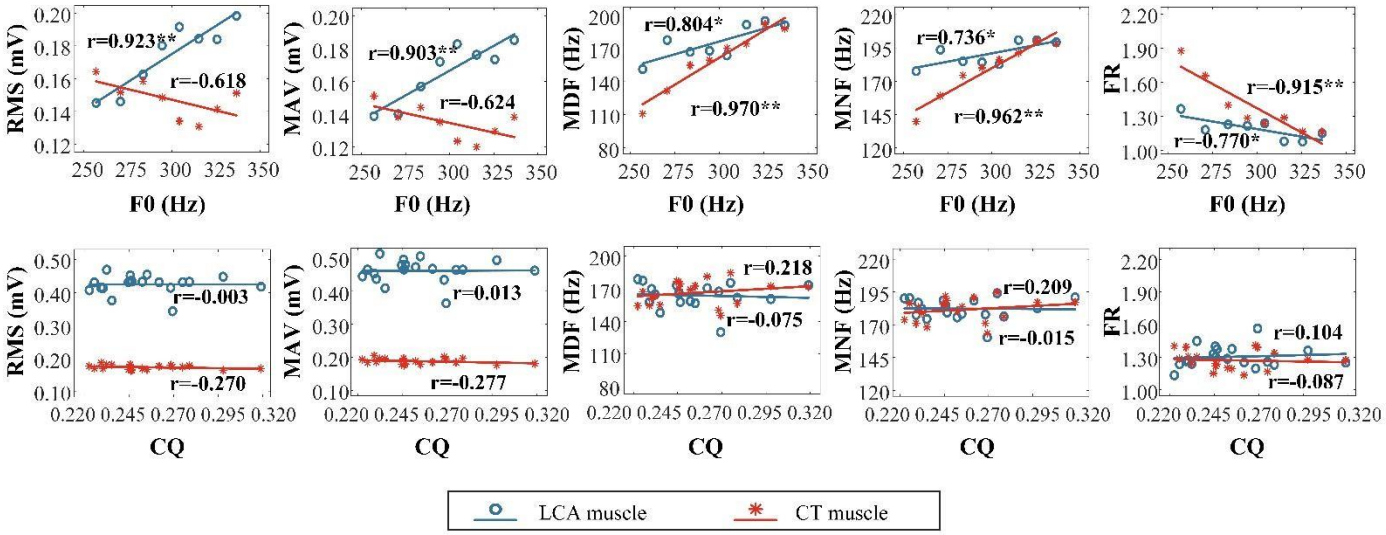


Fig. 2. The relationships between sEMG features and acoustic features. In each subfigure, the scatter plots show extracted sEMG features of LCA (blue color) and CT (red color) muscles, while the solid lines represent linear fitting curves of the scatter data with marked Pearson's correlation coefficients between sEMG features and CQ (* denotes $p < 0.05$ and ** denotes $p < 0.01$).

Conclusions:

Surface electromyography is feasible and effective to non-invasively extract and quantitatively evaluate dynamic activity of CT and LCA muscle during phonation, which is promising to develop into a practical technique for in-vivo human research on neuromuscular control and clinical diagnosis of laryngeal disorders.

Acknowledgements:

This work is supported by the National Natural Science Foundation of China under Grant 12274340.

References:

[1] V. J. Boucher, C. Ahmarani, and T. Ayad, "Physiologic features of vocal fatigue: Electromyographic spectralcompression in laryngeal muscles," Laryngoscope, vol. 116, no. 6, pp. 959–965, Jun. 2006.

[2] J. Mart´ınez, O. Valencia, and M. Za˜nartu. (2023, Mar.). High-density electromyography as a method for estimating laryngeal muscle activity: a preliminary study. Presented at Proc. SPIE. Int. Soc. Opt. Eng. [Online]. Available: http://dx.doi.org/10.1117/12.2670139.

[3] G. Luegmair, D. K. Chhetri, and Z. Zhang. (2014). The role of thyroarytenoid muscles in regulating glottal closure in an in vivo canine larynx model. Presented at Proc. Mtgs. Acoust. ASA. [Online]. Available: http://dx.doi.org/10.1121/2.0001504.

[4] A. Blitzer et al., "Recommendations of the neurolaryngology study group on laryngeal electromyography," Otolaryngol. Head Neck Surg., vol. 140, no. 6, pp. 782–793, Jun. 2009.

[5] D. K. Chhetri and S. J. Park, "Interactions of subglottal pressure and neuromuscular activation on fundamental frequency and intensity," Laryngoscope, vol. 126, no. 5, pp. 1123–1130, Mar. 2016.

INVESTIGATING THE CORRELATION BETWEEN DIVERSE VOCAL TASKS AND INTERMUSCULAR COHERENCE: A PILOT STUDY USING HIGH-DENSITY ELECTROMYOGRAPHY

Josué Martinez[1,2], Loreto Rojas[3], Matías Zañartu [1,2]

[1] Department of Electronic Engineering, Universidad Técnica Federico Santa Maria, Valparaíso, Chile.

[2] Advanced Center for Electrical and Electronic Engineering, Universidad Técnica Federico Santa Maria, Valparaíso, Chile

[3] Electrical Engineering Department, University of Concepción, Concepción, Chile

***Keywords*: High-density electromyography, intermuscular coherence, muscle tension dysphonia.**

Abstract:

Objectives / Introduction: Excessive voice use and compensatory intrinsic and extrinsic laryngeal muscle tension are considered potential factor affecting several voice disorders such as muscle tension dysphonia, Parkinson's disease, laryngeal dystonia, among others. Laryngeal electromyography is a diagnostic technique employing hook-wire intramuscular electrodes that is currently limited to the examination of specific neurological conditions due to its highly invasive nature. Surface electromyography (sEMG) is a non-invasive approach that has the potential to capture extrinsic laryngeal tension that may exhibit altered activity in subjects with dysphonia [1], but has yielded conflicting results, likely due to the variability in electrode configuration, phonatory tasks, outcome measures, and normalization procedures. High-density sEMG (two-dimensional arrays) has the potential to address some of these limitations by extending the contact surface, increasing the number of channels, and providing means to better decompose and analyze the sEMG signals [2]. Furthermore, intermuscular coherence networks have recently been employed using traditional bipolar electrodes to comprehensively investigate the muscular system during various phonatory tasks and have exhibited a more consistent behavior to describe muscle activity than spectrally-temporal features such as Root Mean Square (RMS) and Power Spectral Density (PSD) [3][4].

The aim of this work is to utilize high-density sEMG and intermuscular coherence as a non-invasive method to study laryngeal muscle activity and correlate it with vocal tasks, benefiting from an increased number of acquisition channels. It is anticipated that the efficacy of intermuscular coherence as an estimator of laryngeal muscle tension will be determined and the advantages of high-density sEMG recording will be explored, in comparison with traditional electromyography processing techniques.

Methods: Muscle activity was recorded using an electromyography device (Quattrocentro, OT Bioelettronica, Torino, Italy) in monopolar mode with a sampling frequency of 2000 Hz. A 64-channel (5 x 13) electrode grid with a distance of 8 mm between electrodes was used. A series of vocal tasks were conducted, each with varying levels of loudness and pitch conditions, for the sustained vowels /i/, /a/, and the repeated syllable /pae/. In the first task group, three levels of loudness and three pitch ranges were utilized, each produced for 8 seconds for the sustained vowels /i/, /a/ and syllable /pae/. In the second group, an ascending and descending pitch glide was performed for a vowel /i/. The location of the electrode array (64 channels) was oriented considering the trajectory of the sternocleidomastoid muscle (Fig. 1A). Two subjects, one male and one female, with no history of vocal pathology completed the vocal tasks for this initial pilot study, in accordance with the approved IRB protocol of the FONDECYT 1230828 project.

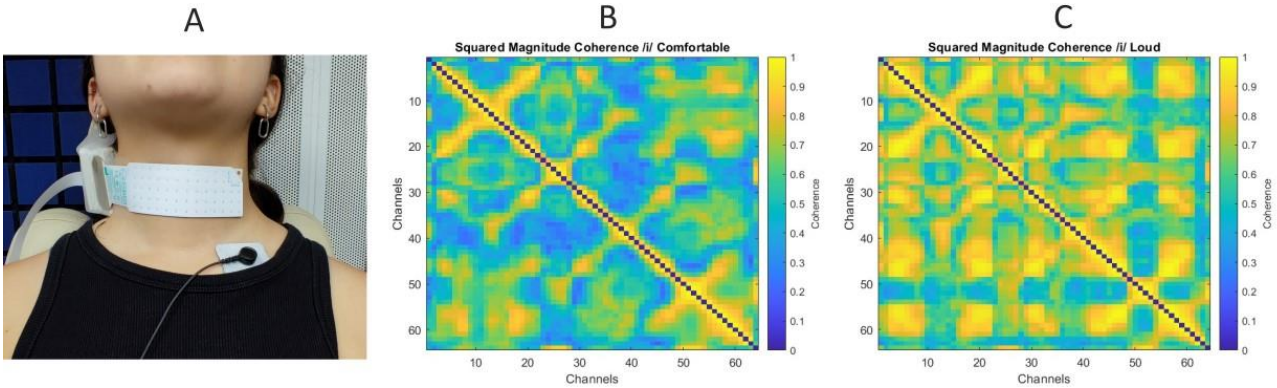


Fig 1. A. Location of the matrix considering different anatomic points. **B.** Matrix of squared magnitude coherence for the vowel /i/ at comfortable amplitude. **C.** Matrix of squared magnitude coherence for the vowel /i/ at loud amplitude

Magnitude squared coherence was calculated using Welch power spectral density with a Hamming window of 2048 samples and 50% overlap. The maximum coherence component in the 10-100 Hz range was selected for each sensor pair. Utilizing the maximum coherence value obtained for each channel, the coherence matrix was constructed, followed by the computation of the average coherence across all channels (Fig. 1B,C). To facilitate a comparison between muscle coherence and traditional methods, RMS value was calculated.

Results: The results illustrate that the concurrent utilization of high-density sEMG and intermuscular coherence has the potential to differentiate both changes in vocal parameters (namely, volume and pitch) and various vocal tasks. This distinction is attributed to the increased number of channels capable of capturing subtle shifts associated with vocal production. This is initially evidenced by the visualization of the coherence matrix, revealing electrode channels with heightened coherence in tasks involving high pitch or loudness. A mean normalized coherence value of 0.81 was obtained for a loud amplitude and a value of 0.53 for a comfortable amplitude of the vowel /i/ (Table 1). Additionally, a symmetry in the coherence matrix is noticeable, indicating a certain uniformity in the distribution of extrinsic laryngeal muscle activity. In this pilot study the ability of intermuscular coherence to distinguish between vocal parameters and tasks proved to be significantly superior to conventional metrics such as RMS, which did not yield satisfactory outcomes, exhibiting very similar patterns across different phonatory tasks.

Table 1. Mean values of the preliminary results.

Task	Pitch	Amplitude	RMS	Coherence
/a/	Habitual	Comfortable	0.12	0.53
		Loud	0.15	0.81
/i/	Habitual	Soft	0.11	0.50
		Comfortable	0.11	0.56
		Loud	0.14	0.68
/pae/	Habitual	Soft	0.20	0.66
		Comfortable	0.23	0.71
		Loud	0.30	0.85
/i/	High	Comfortable	0.14	0.57
	low		0.25	0.73
/i/	Pitch glide ascending	Comfortable	0.42	0.79
	Pitch glide descending		0.37	0.65

Conclusions: This study illustrates potential and sensitivity of intermuscular coherence analysis for reflecting subtle modulations in vocal output, particularly in detecting changes in vocal parameters. The preliminary results suggest that the high functional synchrony of cervical muscles produces a pronounced coherence response during loud pitch exercises, advocating that laryngeal performance can be assessed through intermuscular coherence. The use of highdensity sEMG presents a more comfortable and convenient approach, requiring the placement of only a single electrode instead of multiple electrodes across different areas. The augmented number of acquisition channels also facilitates the monitoring of a broader area, enhancing the detection of finer muscle activity variations. Future efforts include extending this methodology to a broader array of phonatory tasks and a larger patient and control cohort for the acquisition of comprehensive statistical metrics.

Acknowledgements:

This research was supported by the National Institutes of Health (NIH) National Institute on Deafness and Other Communication Disorders grant P50 DC015446 and by ANID grants BASAL FB0008 and FONDECYT 1230828. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

References:

[1] I. Hocevar-Boltezar, M. Janko, and M. Zargi, “Role of surface EMG in diagnostics and treatment of muscle tension dysphonia,” Acta Otolaryngol., vol. 118, no. 5, pp. 739–743, Sept. 1998.

[2] Merletti, R., & Muceli, S. (2019). Tutorial. Surface EMG detection in space and time: Best practices. Journal of Electromyography and Kinesiology, 49, 102363.

[3] O’Keeffe, R., Shirazi, S. Y., Mehrdad, S., Crosby, T., Johnson, A. M., & Atashzar, S. F. (2021). Characterization of cervical-cranial muscle network in correlation with vocal features. bioRxiv, 2021-09.

[4] O'Keeffe, R., Shirazi, S. Y., Mehrdad, S., Crosby, T., Johnson, A. M., & Atashzar, S. F. (2023, April). Perilaryngeal functional muscle network in patients with vocal hyperfunction-a case study. In 2023 11th International IEEE/EMBS Conference on Neural Engineering (NER) (pp. 1-5). IEEE.

Dynamic 3D Magnetic Resonance Imaging of the Vocal Fold Oscillation

Johannes Fischer[1], Louisa Traser[2,3], Bernhard Richter[2,3], Michael Bock[1]

[1] Dept. of Radiology, Medical Physics, Medical Center University of Freiburg, Faculty of Medicine, University of Freiburg, Freiburg, Germany

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

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

***Keywords*: Vocal folds, high-speed imaging, Magnetic Resonance Imaging**

Abstract:

Objectives / Introduction:

Various techniques have been developed to measure vocal fold (VF) oscillations [1]–[5]. However, due to the restricted anatomical access in the larynx, these methods only measure particular aspects such as the caudal surface [3], the layered structure [4] or the tissue velocity [5]. A 3D, temporally resolved measurement of the VF oscillation will increase our understanding of the origin of voice and may guide therapeutic decisions, e.g. when tumor resection is required [6]. We have previously shown that magnetic resonance imaging (MRI) is capable of achieving high spatio-temporal resolution to image the VF oscillation in 2D [7], [8]. In our previous work, we have used rapidly switched magnetic field gradients to encode the tissue motion. In this work, we report on a new approach that minimizes gradient switching and allows for isotropic encoding in all 3 spatial dimensions, with sub-millimeter isotropic spatial and sub-millisecond temporal resolution.

Methods:

MR-data was acquired on a clinical 3T MRI (Siemens Prisma) using a zero echo time (ZTE) pulse sequence. With ZTE the duration of the simultaneous encoding and acquisition of the MR-signal was 0.64ms. To acquire an image with isotropic spatial resolution, the direction of the encoding gradient (G = 23.6mT/m) was changed after each acquisition [9] and data acquisition was repeated every 0.85ms. The MR-signal is acquired with a small circular coil (d=7cm) at the larynx of the volunteer. To acquire an image with 0.8mm spatial resolution, data acquisition was repeated 287.200 times and required 4min 20s in total. The male volunteer (26y, 80kg) was lying in a supine position and asked to sing at a constant frequency of (f = 150Hz) with intermittent breathing at will. Shifts in larynx position over the measurement can cause image blurring. To compensate for these shifts, one-dimensional MR navigator data were acquired every 255ms in superior-inferior (S/I) direction and every 510ms in right-left (L/R) and anterior-posterior (A/P) direction. This allows for retrospective tracking of the larynx position relative to a reference position. Variations in larynx position are then corrected using Fourier shift theorem.

To reconstruct the VF oscillation, MR-data were sorted according to the motion phase during signal acquisition. Therefore, a microphone was placed at the mouth of the volunteer to record the acoustic signal. To suppress the noise from the MRI machine, a low-pass Butterworth filter (6^{th} order, f_0 = 180Hz) was applied to the acoustic signal. The VF motion phase was then obtained from a fit of a sine function for each acquisition. Using this phase, data are sorted into 10 phases (frames).

Results:

The mean phonation frequency was 149±2Hz, yielding a temporal resolution of 0.67ms in each of the 10 frames (Figure 1). Transverse images of the larynx show a wide opening of up to 2.4mm across the glottis, and the coronal view shows a vertical displacement of 1mm during the reconstructed oscillation cycle. From the images in sagittal orientation we could estimate the variation of contact area between 46 and 78mm^2 in the volunteer.

During phonation the maximal displacements of the larynx in S/I direction were –4mm and +3mm. The standard deviation over all estimated larynx displacements was □□= 0.8mm. Via thresholding, we extracted surface data of the glottis (Figure 2).

Conclusions:

In this work we for the first-time measure VF oscillations in humans with a full high spatio-temporal 3D+t resolution. The ZTE pulse sequence is advantageous as it provides a very short encoding time and low acoustic noise compared to conventional pulse sequences, making it the ideal imaging method during speech and singing.

A disadvantage of ZTE is the low tissue contrast, which currently does not provide a differentiation between the layers of the VF. However, for VF motion studies the boundary between tissue and the air in the glottis can be visualized. In our previous work, instead of the MR-compatible microphone, we have used an electroglottogram (EGG) for signal synchronization. While the optical microphone improves measurement setup times and patient comfort, the reconstructed images can no longer be correlated to features in the EGG signal.

A higher spatio-temporal resolution is currently limited by the spatial selectivity of the MR excitation pulse. Solutions to this problem have been proposed and will be applied in a next step [10]. In voice research, the access to the vocal folds is limited, requires unpleasant insertion of a laryngoscope into the pharynx and can be prone to perspective distortion [11]. Computational simulations and ex-vivo studies showed that vocal control is three-dimensional [12]. Thus in the future, we want to apply this new technique in a volunteer study to visualize and quantify medial surface vertical thickness of the VF during vibration in vivo and asses its role in regulating the glottal closure pattern for different voice qualities.

Acknowledgements:

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

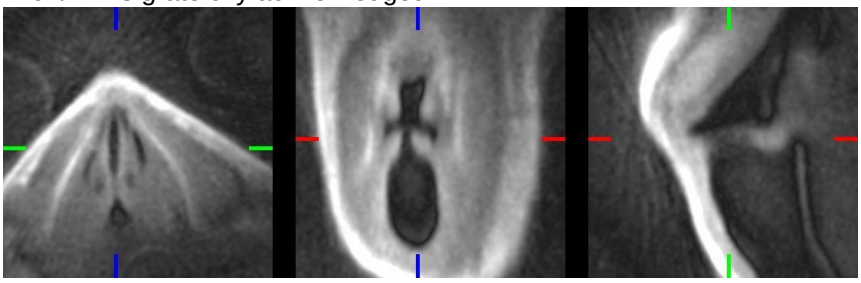


Figure 1: Transverse, coronal and sagittal views of the vocal folds. Images are taken from the 3D+t MR data. White lines indicate the position of the other 2 views. Animated version.

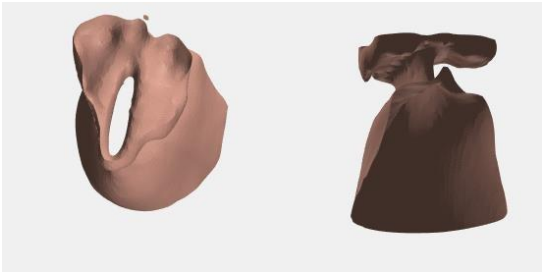


Figure 2: Reconstruction of the surface area in the glottis during phonation. Left: view from top (artificial laryngeal stroboscopy), Right: Rotating view from below the vocal fold plane. Animated version.

References:

[1] M. J. Oertel, “Laryngostroboscopische Beobachtungen über die Bildung der Register bei der menschlichen Stimme,” *Centralbl Med Wiss*, vol. 16, pp. 99–101, 1878.

[2] J. G. Švec and H. K. Schutte, “Videokymography: High-speed line scanning of vocal fold vibration,” *Journal of Voice*, vol. 10, no. 2, pp. 201–205, Jan. 1996, doi: 10.1016/S0892-1997(96)80047-6.

[3] G. Luegmair, D. D. Mehta, J. B. Kobler, and M. Döllinger, “Three-Dimensional Optical Reconstruction of Vocal Fold Kinematics Using High-Speed Video With a Laser Projection System,” *IEEE Transactions on Medical Imaging*, vol. 34, no. 12, pp. 2572–2582, Dec. 2015, doi: 10.1109/TMI.2015.2445921.

[4] C. A. Coughlan *et al.*, “In vivo cross-sectional imaging of the phonating larynx using long-range Doppler optical coherence tomography,” *Sci Rep*, vol. 6, no. 1, Art. no. 1, Mar. 2016, doi: 10.1038/srep22792.

[5] Y. W. Shau, C. L. Wang, F. J. Hsieh, and T. Y. Hsiao, “Noninvasive assessment of vocal fold mucosal wave velocity using color doppler imaging,” *Ultrasound Med Biol*, vol. 27, no. 11, Art. no. 11, Nov. 2001, doi: 10.1016/s0301-5629(01)00453-7.

[6] P. Schultz, “Vocal fold cancer,” *European Annals of Otorhinolaryngology, Head and Neck Diseases*, vol. 128, no. 6, Art. no. 6, Dec. 2011, doi: 10.1016/j.anorl.2011.04.004.

[7] J. Fischer, T. Abels, A. C. Özen, M. Echternach, B. Richter, and M. Bock, “Magnetic resonance imaging of the vocal fold oscillations with sub-millisecond temporal resolution,” *Magnetic Resonance in Medicine*, vol. 83, no. 2, Art. no. 2, 2020, doi: 10.1002/mrm.27982.

[8] J. Fischer *et al.*, “Sub-millisecond 2D MRI of the vocal fold oscillation using single-point imaging with rapid encoding,” *Magn Reson Mater Phy*, Sep. 2021, doi: 10.1007/s10334-021-00959-4.

[9] D. Piccini, A. Littmann, S. Nielles-Vallespin, and M. O. Zenge, “Spiral phyllotaxis: The natural way to construct a 3D radial trajectory in MRI,” *Magnetic Resonance in Medicine*, vol. 66, no. 4, pp. 1049–1056, 2011, doi:
10.1002/mrm.22898.

[10] C. Li, J. F. Magland, A. C. Seifert, and F. W. Wehrli, “Correction of Excitation Profile in Zero Echo Time (ZTE) Imaging Using Quadratic Phase-Modulated RF Pulse Excitation and Iterative Reconstruction,” *IEEE Transactions on Medical Imaging*, vol. 33, no. 4, pp. 961–969, Apr. 2014, doi: 10.1109/TMI.2014.2300500.

[11] R. Veltrup, S. Kniesburges, and M. Semmler, “Influence of Perspective Distortion in Laryngoscopy,” *Journal of Speech, Language, and Hearing Research*, vol. 66, no. 9, pp. 3276–3289, Sep. 2023, doi: 10.1044/2023_JSLHR23-00027.

[12] Z. Zhang, “Vocal Fold Vertical Thickness in Human Voice Production and Control: A Review,” *Journal of Voice*, Mar. 2023, doi: 10.1016/j.jvoice.2023.02.021.

EVALUATION OF THE FREQUENCY-DEPENDENT YOUNG'S MODULUS OF THE VOCAL FOLDS BY OPTICAL COHERENCE TOMOGRAPHY IN COMBINATION WITH PIPETTE ASPIRATION

Raphael Lamprecht[1], Florian Scheible[1], Alexander Sutor[1]
[1]Institute of Measurement and Sensor Technology,
UMIT TIROL – Private University for Health Sciences and Health Technology, Hall in Tirol, Austria

***Keywords*: vocal folds, frequency dependent material properties, optical coherence tomography, pipette aspiration**

Abstract:

Objectives / Introduction:
The vocal folds produce the primary sound of phonation. Therefore, it is important to consider any changes in material characteristics that may occur due to vibration frequency when analyzing vocal production. Various techniques for determining viscoelastic properties have been developed, such as the tensile test, rheometer, atomic force microscopy as well as indentation measurements, pipette aspiration, ultrasound elastography and Brillouin microspectroscopy, the last four of which have the potential to be used in vivo.
The microstructure of the superficial layers of the vocal folds could be effectively visualized with optical coherence tomography (OCT) [1]. The pipette aspiration technique, on the other hand, has shown that it is capable of measuring frequency-dependent properties [2]. Published data suggest that the stiffness of the vocal fold tissue increases with frequency [3, 4].
The combination of these two techniques can provide a comprehensive understanding of the tissue's structure as well as its viscoelastic properties.
This abstract presents the OCT pipette aspiration as a novel measurement technique in voice science and its results on the frequency-dependent elastic properties of the superficial layers of the vocal folds.

Methods:
An OCT (OQ Labscobe 2.0/X, Lumedica) is used to record the B-mode images, with a 3D printed pipette mounted to the sensor head. The pipette has an inner radius of 2.5 mm, an outer radius of 3.5 mm and is shielded from the optical system by a sapphire window. The pipette is pressed vertically onto the tissue surface by a linear drive, whereby a contact pressure of 0.05 N is desired. The tissue is excited by a customary loudspeaker and simultaneously the applied pressure is measured by a pressure transducer (XCS-093±5SG, Kulite). The measurement setup is shown in Figure 1a.
The excitation frequency range is between 1 Hz and 100 Hz, with 200 OCT images and 4 seconds of pressure data recorded for each of the 34 steps.
The pressure fluctuation follows a sinusoidal pattern; therefore, a vibration cycle can be reconstructed using the pressure data as a reference. This step is necessary due to the low sample rate (approx. 30 Hz) of the OCT. Furthermore, the surface of the tissue is detected using a second-degree polynomial fit to the brightest line in recorded B-scan. Using the maximum displacement of the specimen surface Δdd and the maximum pressure fluctuation Δpp, the Young's modulus EE^{PPPP} of a Neo-Hookean material can be estimated by

$$E^{PA} = \frac{a\,\Delta p}{1.07\,\Delta d},$$

where aa is the inner radius of the pipette [5]. In addition, the static Young's modulus EE^{II} is measured using the pipette for an indentation measurement [6].
On a thawed porcine larynx, the pipette is placed in the middle of the vocal folds and four iterations of the entire frequency range are measured in each case.

Results:

Figure 1b shows an exemplary OCT image in which the tissue surface and the layered structure of the tissue are clearly visible.

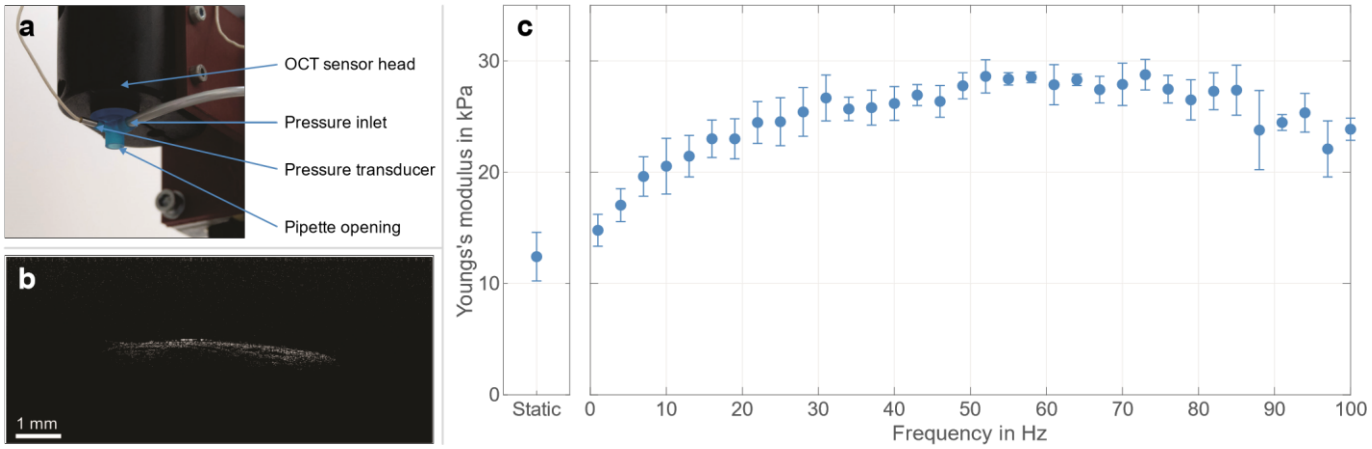


*Figure 1: Panel **a** displays the measurement setup. Panel **b** shows an OCT image of the vocal folds. Panel **c** indicates the Young's modulus of the vocal fold tissue. The marker represents the mean value for all iterations (n = 4), and the error bar indicates one standard deviation of the same data set.*

Figure 1c displays the mean values for both the static and pipette measurements for each step. It is observed that the Young's modulus increases with increasing frequency, although there is nearly no change above 50 Hz.

Conclusions:

The data presented indicates an increase in Young's modulus for higher frequencies, consistent with previously published data. However, based on the measurements on one larynx with four iterations, no statistically significant statement can be made.

Nevertheless, it is shown that the combination of pipette aspiration and optical coherence tomography is capable of measuring the frequency-dependent Young's modulus of the vocal folds. Additionally, it is possible to record B-mode images of the vocal fold structure simultaneously. Therefore, the presented method expands the diagnostic possibilities of non-destructive tissue characterization and may be beneficial in a clinical setting.

References:

[1] M. J. Hawkshaw, J. B. Sataloff, and R. T. Sataloff, “New concepts in vocal fold imaging: a review,” *J. Voice*, vol. 27, no. 6, pp. 738–43, Jan. 2013, doi: 10.1016/j.jvoice.2013.05.011.

[2] F. Scheible *et al.*, “Behind the Complex Interplay of Phonation: Investigating Elasticity of Vocal Folds With Pipette Aspiration Technique During Ex Vivo Phonation Experiments,” *J. Voice*, p. S0892199723000966, Mar. 2023, doi: 10.1016/j.jvoice.2023.03.001.

[3] R. W. Chan and M. L. Rodriguez, “A simple-shear rheometer for linear viscoelastic characterization of vocal fold tissues at phonatory frequencies,” *J. Acoust. Soc. Am.*, vol. 124, no. 2, pp. 1207–1219, Aug. 2008, doi: 10.1121/1.2946715.

[4] S. S. Teller *et al.*, “High-Frequency Viscoelastic Shear Properties of Vocal Fold Tissues: Implications for Vocal Fold Tissue Engineering,” *Tissue Eng. Part A*, vol. 18, no. 19–20, pp. 2008–2019, Oct. 2012, doi: 10.1089/ten.tea.2012.0023.

[5] M.-G. Zhang, Y.-P. Cao, G.-Y. Li, and X.-Q. Feng, “Pipette aspiration of hyperelastic compliant materials: Theoretical analysis, simulations and experiments,” *J. Mech. Phys. Solids*, vol. 68, pp. 179–196, Jan. 2014, doi: 10.1016/j.jmps.2014.03.012.

[6] E. Willert, Q. Li, and V. L. Popov, “The JKR-adhesive normal contact problem of axisymmetric rigid punches with a flat annular shape or concave profiles,” *Facta Univ. Ser. Mech. Eng.*, vol. 14, no. 3, p. 281, Jan. 2016, doi: 10.22190/FUME1603281W.

VOCDOC –
A NOVEL APP TO CAPTURE LONG-TERM VOICE VARIATIONS IN THE WILD

Florian B. Pokorny[1,2], Claus Gerstenberger[1], Julian Linke[3], Florian Eyben[4], Björn W. Schuller[2,4], Markus Gugatschka[1]

[1] Division of Phoniatrics, Medical University of Graz, Graz, Austria
[2] CHI – Chair of Health Informatics, Technical University of Munich, Munich, Germany
[3] Signal Processing and Speech Communication Laboratory, Graz University of Technology, Graz, Austria [4] audEERING GmbH, Gilching, Germany

***Keywords*: Vocal Fatigue; Voice Features; Mobile Application; Digital Health**

Abstract:

Objectives / Introduction: A healthy voice is a fundamental 'tool' for social and occupational life. A number of professions, e.g., teacher, lecturer, or kindergarten worker, come along with excessive voice use. Over years, prolonged vocal loading can lead to substantial voice problems, such as chronic hoarseness. In serious cases, social isolation and occupational disability can be the consequences. An early identification and characterisation of vocal dysfunction are thus essential as they enable a targeted intervention. However, standard outpatient voice assessments only provide a snapshot of a patient's voice and usually do not capture effects of vocal loading that emerge after a certain longer period of voice use in everyday life settings. Therefore, we suggest a mobile solution for longitudinally collecting objective voice parameters 'in the wild' and making them clinically accessible.

In a preliminary study (Pokorny et al., 2024), we conducted voice recordings of four voice healthy speakers, who completed a 90-minute reading task. Acoustic analyses of an extended set of extracted voice parameters revealed individual effects of vocal fatigue across different speakers. Only a few parameters turned out to universally describe vocal changes over time.

In this work we wanted to take an important step towards practical implementation. We aimed to develop a prototype of a long-term voice recording and voice parameter extraction app – the *VocDoc*, and demonstrate its basic functionality under real-world conditions.

Methods: We implemented the *VocDoc* prototype for iPhone. For constant recording quality, a clip-on microphone is expected to be attached to the collar and plugged to the phone. The *VocDoc* was supplied with an intelligent and noiserobust voice activity detection input stage, i.e., a machine learning-based agent that initially learns intrinsic voice characteristics of a specific user based on a few sample utterances and, thenceforth, automatically detects whenever the user is speaking. For each detected speech segment of the user, the *VocDoc* finally extracts a set of 84 voice parameters, such as statistical functionals of the fundamental frequency, voicing probability, alpha ratio, Hammarberg index, or spectral flux, in real time and stores them locally in form of a vector together with time stamps in a text file. For data privacy reasons, the *VocDoc* does not store raw audio. Feature extraction is realised by means of the widely-used open-source toolkit openSMILE (Eyben et al., 2010).

For a proof-of-concept evaluation, we equipped a 32-year-old male university teacher with the *VocDoc* and collected data while he was giving a 90-minute lecture.

Results: During the 90-minute test run under realistic conditions, the *VocDoc* detected 7836 speech segments of the target speaker, successfully extracted the segment-wise voice parameters, and stored them in a text file. The trajectories of several voice parameters were subject to high fluctuations. Nevertheless, we could observe overall linear trends over time, such as an increase of the mean fundamental frequency or the standard deviation of voicing probability in voiced regions; see Figure 1.

Conclusions: We introduced and initially tested a novel app – the *VocDoc* – as an easy-to-apply mobile framework for the long-term recording of clinically relevant voice parameters of a patient in everyday life settings. Collected data are intended to be visualised and interpreted by phoniatricians and speech-language therapists for diagnostic reasoning, intervention planning, and therapy success evaluation. Next steps comprise the collection and analysis of patient voice data by means of the *VocDoc*, a final selection of supported voice parameters in consultation with healthcare professionals, as well as an integration of machine learning-based features, such as an automatic detection of pathological voice segments.

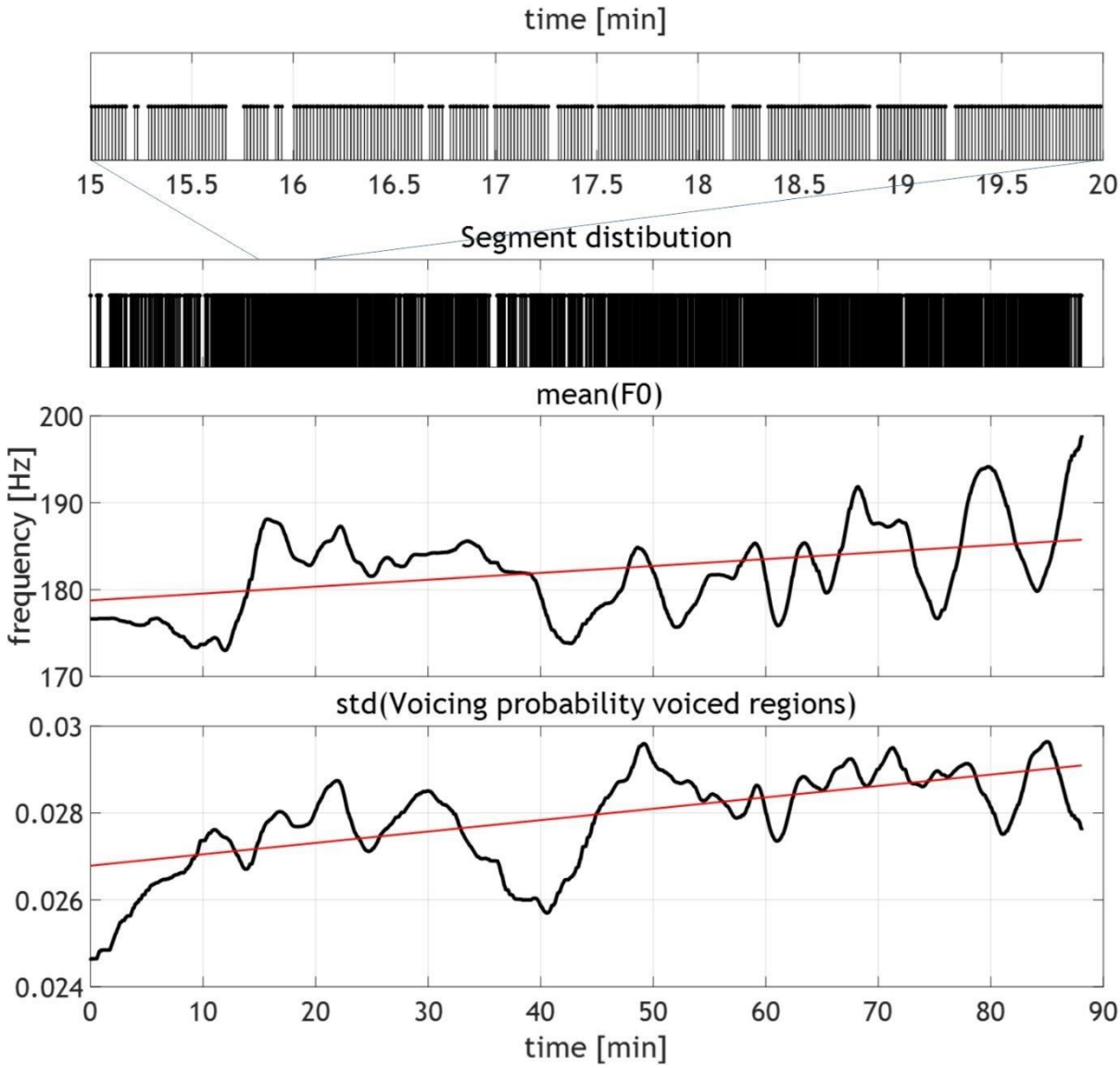


Figure1: Distribution of detected speech segments alongside a 90-minute lecture and smoothed trajectories of two segment-wisely extracted example voice parameters including linear fitting line (red).

References:

Eyben F, Wöllmer M, Schuller B. openSMILE: The Munich versatile and fast open-source audio feature extractor. Proceedings of the ACM International Conference on Multimedia, 1459–1462, 2010.

Pokorny FB, Linke J, Seddiki N, Lohrmann S, Gerstenberger C, Haspl K, Feiner M, Eyben F, Hagmüller M, Schuppler B, Kubin G, Gugatschka, M. VocDoc, what happened to my voice? Towards automatically capturing vocal fatigue in the wild. Biomedical Signal Processing and Control, 88, 105595, 2024.

EFFECT OF THYROPLASTY TYPE I IMPLANT LOCATION ON GLOTTAL MEDIAL SHAPE DURING PHONATION OF A CANINE LARYNX MODEL

Charles Farbos de Luzan[1], Jacob Michaud-Dorko[1], Rebecca Howell[1], Ohad Cohen[1], Gregory Dion[1], Ephraim Gutmark[2], Liran Oren[1]

[1] Department of Otolaryngology, University of Cincinnati, Cincinnati, OH, USA
[2] Department of Aerospace Engineering, University of Cincinnati, Cincinnati, OH, USA

***Keywords*: Voice; Surgery; Digital Image Correlation; Highspeed**

Abstract:

Objectives / Introduction: Common operations for unilateral vocal fold paralysis (UVFP), such as Thyroplasty Type 1, improve vocal efficiency (VE) by closing the membranous gap between the normal and paralyzed fold. VE measures the degree to which subglottal aerodynamic power is translated to acoustic power. However, even after surgical treatment, some patients report symptoms such as vocal fatigue, decreased projection, and decreased intelligibility in noisy environments. These recalcitrant symptoms will be associated with reduced VE relative to normal voice, suggesting that enhancing VE could further ameliorate these symptoms.

Previous research in the excised canine larynx suggests that this further improvement is possible by certain modifications to the procedure. For example, it was found that medialization of the tissue below the fold (infraglottal region) results in higher VE than medialization of the fold itself (glottal) even though both operations close the mediallateral membranous gap[1,2]. However, translating these surgical findings in animals to patients requires a deeper understanding of how and why various treatment modifications affect VE.

This study's goal is to quantify the effect of implant location on the stiffness gradient, maximum divergence angle, vocal fold vertical height, and vocal efficiency (VE) using an excised canine larynx model.

Methods: 4 larynges were excised from healthy canines, and a hemilaryngectomy was performed to remove the right fold, leaving the trachea intact. The hemilarynx was attached to a transparent Plexiglas plate using a suture at the anterior commissure and two on the thyroid cartilage in order to secure it in place. Sealing was ensured by a bead of silicon sealant, delineating the contact between the thyroid cartilage and the transparent plate. The model was mounted on an aerodynamic nozzle to supply the glottis with conditioned airflow at a known flow rate and pressure. The remaining arytenoid cartilage was positioned using a three-axis prong system. This way, different measurable adduction levels were achievable. Three conditions were evaluated: without an implant, with an implant at the glottal level, and with an implant placed infra-glottal (Fig. 1).

A stereoscopic digital image correlation setup acquired mid-membranous tissue deformation during phonation, using two high-speed cameras and a volumetric laser to illuminate the sample. The three-dimensional medial shape of the glottis was acquired in its pre-phonatory state and then during phonation. A microphone was used to measure the sound pressure level to derive vocal efficiency.

After phonation trials, the bulk stiffness of the tested fold was measured using a micro-indenter.

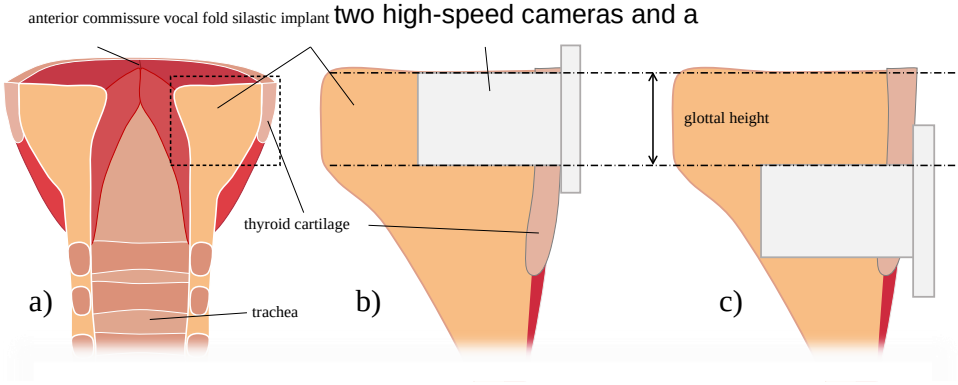


Figure 1 Implant Locations in the Larynx. (a) Cross-section of the mid-coronal plane of larynx post removal of supraglottal tissues, showing internal structure. (b) and (c) are enlarged views from the dashed rectangle in (a), depicting silastic implant placement

Results: The implant location (inferior vs. superior) directly impacted the medial shape and the vibration amplitude. The medialization effect was also noticeable in the prephonatory state (Fig. 2). Trends observed in VE were not consistent across the different larynges.

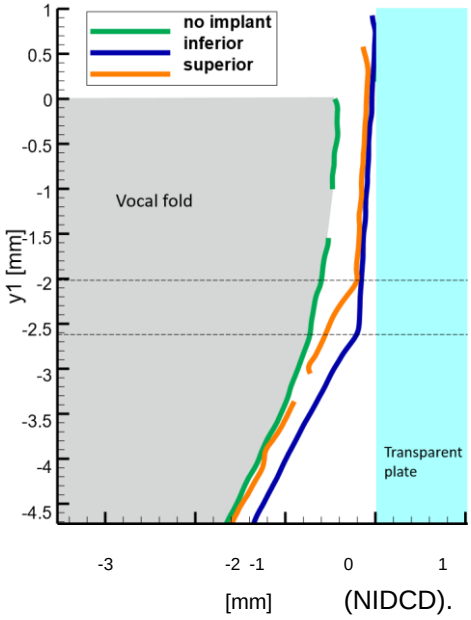


Figure 2 Prephonatory profiles of the vocal fold's shape extracted in the midcoronal plane for the three different implant conditions. Dashed lines emphasize the location of the inferior aspect of the fold.

Conclusions: VE is not so much impacted by vertical height, rather than inferior-superior phase-delay (divergence angle), although implant location affected both. The trends differed from what had been reported in literature previously with full larynges, which could be due to the hemilaryngectomy approach and its associated challenges, such as potential non-anatomically realistic leakages, or vibration onto a hard plate instead of compliant tissue. Statistical analysis is recommended.

Acknowledgements:

The research was funded by NIH Grant No. R01DC009435 from the National Institute on Deafness and Other Communication Disorders (NIDCD).

References:

1. A. Maddox, L. Oren, C. Farbos de Luzan, R. Howell, E. Gutmark, and S. Khosla, "An Ex-vivo Model Examining Acoustics and Aerodynamic Effects Following Medialization With and Without Arytenoid Adduction," Laryngoscope, vol. 133, no. 3, pp. 621–627, Mar. 2023.
2. L. Oren, A. Maddox, C. Farbos de Luzan, C. Xie, G. Dion, E. Gutmark, S. Khosla, "Acoustics and aerodynamic effects following glottal and infraglottal medialization in an excised larynx model," European Archives of Oto-RhinoLaryngology. 2024. In print.

THE INFLAMMATORY AND BIOMECHANICAL EFFECTS OF INHALED CORTICOSTEROIDS ON RABBIT VOCAL FOLD TISSUE

Maya Stevens[1,2], Brendan Olson[3], Anika Isom[4], Douglas Roberts[5], Jenny Pierce[2], Julie M Barkmeier-Kraemer[1,2], Ray Merrill[6], Kristine Tanner[1], Ben Christensen[2,3]

[1] Department of Communication Sciences & Disorders, University of Utah, SLC, UT, USA
[2] Department of Otolaryngology – Head & Neck Surgery, University of Utah School of Medicine, SLC, UT, USA
[3] Department or Surgery, Division of Urology, University of Utah School of Medicine, SLC, UT, USA
[4] Department of Biology, University of Utah, SLC, UT, USA
[5]Department of Bioengineering, University of Utah, SLC, UT, USA
[6]Department of Communication Sciences & Disorders, Brigham Young University, Provo, UT, USA

***Keywords*: Asthma; Dysphonia; Inflammation; Elastic Modulus**

Abstract:

Objectives / Introduction: 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 patients[1]. This association has been attributed to the fact that the drugs must pass through the airway before reaching the lungs. 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: Twenty 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 -80ºC. 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.

Results: 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.

Conclusions: 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.

Acknowledgements:

This work was funded by an NIH R01 to Kristine Tanner (R01 DC016269.

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

INVESTIGATING THE INFLUENCE OF THE VOCAL TRACT ACOUSTO-MECHANICAL RESONANCE ON THE DYNAMICS OF SEMI-OCCLUDED VOCAL TRACT EXERCISES INVOLVING TUBES IMMERSED IN WATER

Andrey Ricardo da Silva[1], Igor A. M. Kawamura[1], Ana Carolina A. M. Ghirardi[2],

[1] Dept. of Mechanical Engineering, Federal University of Santa Catarina, Florianópolis, SC, Brazil
[2] Dept. of Speech-Language Pathology and Audiology, Federal University of Santa Catarina, Florianópolis, SC, Brazil

***Keywords*: SOVTE; Vocal Tract Mechanical Resonance; Resonance Tube**

Abstract:

Introduction/Objectives: Semi-Occluded Vocal Tract Exercises (SOVTE) elevate vocal tract pressure during phonation by increasing air flow resistance. The augmented pressure expands the vocal tract, compelling the speaker to actively engage the muscles around the epilarynx in order to achieve an optimal acoustic impedance match between the vocal folds and vocal tract, thereby producing voice. One popular SOVTE method involves phonating into a tube with one of its terminations submerged in water. The advantages of this technique include a consistent flow resistance, primarily dictated by the tube's immersion depth, and an oscillatory back pressure component arising from bubble release, which is assumed to alleviate muscle tension. This study seeks at understanding the impact of the vocal tract compliance on the bubble release mechanism and the oscillatory component of the back pressure.

Methods: The study is achieved through experimental assessments and comparisons of BRF and vocal tract pressure as functions of flow rate, using tubes connected to both a non-compliant (synthetic) vocal tract and real vocal tracts of six adult subjects.

Results: The comparison reveals significant divergences between the results obtained from non-compliant and real vocal tracts. Real vocal tracts display notably higher pressure fluctuations and the BRF appears to be driven by the vocal tract acousto-mechanical resonance.

Conclusions: The substantial increase in the oscillatory component of the pressure within a real vocal tract indicates feedback loop between the pressure source (bubble release) and the resonator (vocal tract). The increase in oral pressure fluctuation caused by the feedback loop implies that subjects must exert an additional effort in order to adjust their vocal tract, potentially enhancing kinesthetic awareness and improving long-term voice production.

Acknowledgements:

The authors would like to thank FINEP/Brazil for grant number 01.16.0044.00 (0346/15)

POSTERS

AUTOMATIC CLASSIFICATION AND DIAGNOSIS OF MUSCLE TENSION DYSPHONIA USING MACHINE LEARNING

CANCELLED.

INVESTIGATING THE SOUND GENERATION IN THE HUMAN VOICE BASED ON PARTICLE IMAGE VELOCIMETRY

Christoph Näger[1], Alexander Lodermeyer[2], Stefan Kniesburges[3], Stefan Becker[1]

[1] Institute of Fluid Mechanics, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Bavaria, Germany
[2] Institute of Process Machinery and Fluid System Dynamics, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Bavaria, Germany
[3] Division of Phoniatrics and Pediatric Audiology at the Department of Otorhinolaryngology, Head and Neck Surgery, University Hospital Erlangen, Medical School at Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Bavaria, Germany

***Keywords*: particle image velocimetry, acoustics, human phonation, dynamic mode decomposition**

Abstract:

Introduction

The human voice is generated in a complex interplay of fluid flow, structural vibration and acoustics. In this process, the vocal folds are stimulated to vibrate by a flow of air $\dot{V}$ from the lungs. This oscillation in turn leads to a modulation of the air flow, resulting in a pulsating jet flow in the vocal tract. The sound that constitutes thereby is mainly generated aeroacoustically from the turbulent pulsating jet flow. This sound is filtered through the vocal tract and radiated through the mouth, resulting in the voice. A detailed investigation of this process in vivo is often hindered due to limited access to the flow field in the larynx and trachea. Therefore, clinicians often rely on measurements of the vocal fold vibration for the diagnosis of speech impairments.

In an academical setting, these limitations are often circumvented by investigating the phonation process on excised human or animal larynges or artificial vocal folds. This allows for a better access of the regions of interest and more indepth analysis of the underlying physics. For performing detailed flow analysis in and above the glottis, particle image velocimetry (PIV) has been proven as a useful tool in the past. Classical planar PIV measurements allowed to study the basic features of the supra- and intraglottal aerodynamics (see e.g. Oren et al. (2014), Lodermeyer et al. (2015)). Also tomographic PIV measurements have been applied in voice research recently, allowing to study volumetric quantities such as the maximum flow declination rate (de Luzan et al. (2020)).

However, especially high-speed PIV has emerged as a valuable method for connecting supraglottal aerodynamics and the resulting acoustics of the voice (Lodermeyer et al. (2021), Näger et al. (2023)). Therefore, in this work we use highspeed PIV measurements to investigate the human phonation process on a synthetic larynx model. For a better understanding of the connection between the flow field and its corresponding acoustic sources, we additionally apply dynamic mode decomposition (DMD) to the measured flow fields, which enables us to gain insight into coherent flow structures at for the phonation process relevant frequencies.

Methods

Synthetic vocal folds were cast from a single layer of silicone. Their shape was based on the M5 model by Scherer et al. (2001). The experimental setup is depicted in figure 1. A vocal tract of rectangular cross section with a length of 180mm was added downstream of the vocal folds. A silencer was added upstream to attenuate emerging sound in the inflow hose. The vocal tract wall was made of glass to allow for optical access for the high-speed camera. PIV measurements were conducted with a measurement frequency of 2x5 kHz and a pulse distance of 4µs. The measured velocity fields were then used in combination with a Poisson solver to obtain pressure fields as well. This allowed us to compute aeroacoustic source terms of the PCWE on the PIV-grid. These could be interpolated onto a finite element grid, which enabled us to do acoustic radiation simulations based on our PIV measurements.

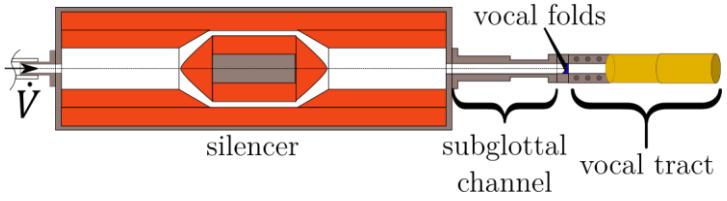


Figure 1 – The experimental setup. The vocal fold position is indicated between the vocal tract and the subglottal channel. A silcencer is placed upstream to attenuate emerging sound in the inflow hose.

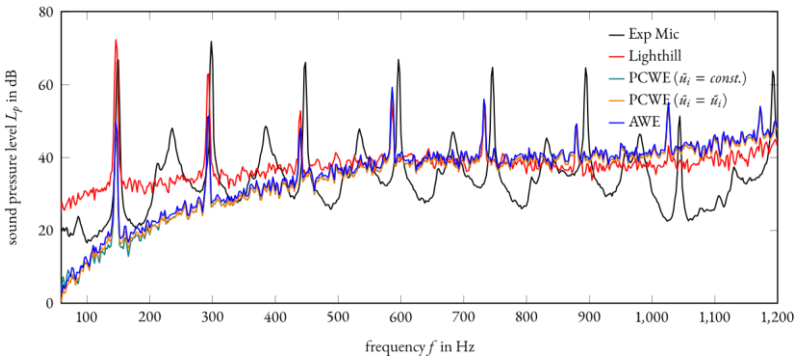


Figure 2 – Comparison of the the sound pressure level Lp for the acoustic analogies based on a perturbation ansatz PCWE, and AWE compared to the experimental microphone measurement (black) and the simulation based on Lighthills´s formulation (red)

Results and Discussion

The aerodynamics in the VT were shown to be dominated by a fundamental oscillation frequency and its corresponding higher harmonics. The DMD analysis revealed that the basic oscillation frequency of the vocal folds is the dominant cause for the oscillation of the jet. The higher harmonic frequencies on the other hand are characterized by periodically shedding vortex pairs of decreasing size with increasing frequency. These vortex pairs lead to strong aeroacoustic sources. A fourier transformation of the source terms revealed, that the main source region for the base frequency was located just downstream of the glottis. For the higher harmonics, the source regions moved slightly further downstream in the channel.

The acoustic radiation simulations showed a good agreement with experimental microphone measurements in their broadband level. Also the peak locations at the oscillation frequency and its higher harmonics are captured well. There exist however some differences in the peak levels between the measurement and the simulations.

Conclusion

The described measurement procedure enabled us to perform detailed investigations in analyzing the human phonation process. The DMD analysis revealed vortex pairs shedding at the glottis that can be used to explain the resulting aeroacoustic source terms. An interesting direction to extend this approach is the investigation of two-way fluid-acousticinteraction. Titze (2008) described, that the acoustic standing waves in the vocal tract can under certain circumstances influence the (sub-)glottal flow field as well as on the vocal fold vibration. This phenomenon will be analyzed with the developed approach in future works.

Acknowledgements:

This work is funded by the German Research Foundation (DFG) through the project
“Tracing the mechanisms that generate tonal content in voiced speech”. Project number: 446965891.

References:

- de Luzan, C.F., Oren, L., Maddox, A., Gutmark, E., Khosla, S.M. (2020). Volume velocity in a canine larynx model using time-resolved tomographic particle image velocimetry. Experiments in Fluids, 61, 63.
- Lodermeyer, A., Becker, S., Döllinger, M., Kniesburges, S. (2015). Phase-locked flow field analysis in a synthetic human larynx model. Experiments in Fluids, 59, 77.
- Lodermeyer, A., Bagheri, E., Kniesburges, S., Näger, C., Probst, J., Döllinger, M., Becker, S. (2021). The mechanisms of harmonic sound generation during phonation: A multi-modal measurement-based approach. The Journal of the Acoustical Society of America, 150, 3485-3499.
- Näger, C., Kniesburges, S., Tur, B., Schoder, S., Becker, S. (2023). An Investigation of Acoustic Back-Coupling in Human Phonation on a Synthetic Larynx Model. Bioengineering, 10, 1343.
- Oren, L., Khosla, S., Gutmark, E. (2014). Intraglottal geometry and velocity measurements in canine larynges. The Journal of the Acoustical Society of America, 135, 380-388.
- Scherer, R.C., Shinwari, D., Witt, K.J.D., Zhang, C., Kucinschi, B.R., Afjeh, A.A. (2001). Intraglottal pressure profiles for a symmetric and oblique glottis with a divergence angle of 10 degrees. The Journal of the Acoustical Society of America, 109, 1616-1630.
- Titze, I.R. (2008). Nonlinear source-filter coupling in phonation: Theory. The Journal of the Acoustical Society of America, 131, 2999-3016.

FACE MASKS AND RISK OF VOICE DISORDER IN UNIVERSITY PROFESSORS

Ana Carolina Ghirardi[1,2], Magda Silva Gomes[2,3,4]

[1] Department of Speech-Language Pathology and Audiology, Federal University of Santa Catarina (UFSC), Florianópolis, SC, Brazil

[2] Graduate Studies Program in Speech-Language Pathology and Audiology Federal University of Santa Catarina (UFSC), Florianópolis, SC, Brazil

[3] Santa Teresa Hospital, Health Department of Santa Catarina, SC, Brazil

[4] São Francisco de Assis Hospital, Santo Amaro da Imperatriz, SC, Brazil

***Keywords*: Voice; Teachers; Voice Disorders; Face Masks**

Abstract:

Objectives / Introduction: The use of protective face masks became necessary when teachers returned to in-person classes amid the COVID-19 pandemic. Teachers are considered a professional category with elevated risk for work-related voice disorders and voice symptoms. Teachers wearing a face mask while teaching may face several challenges related to healthy voice production, and the use of these protective devices, albeit necessary when there is risk of airborne disease transmission, may cause vocal overload[1,2]. Studies show that individuals who wore face masks to work during the pandemic reported more vocal symptoms and difficulties in coordinating breath and speech[3]. One study suggests that surgical and N95 protection masks favor speech intelligibility when compared to simple cloth face masks[4]. Therefore, better understanding the challenges and advantages of the use of face masks may aid in planning actions that will prevent work-related voice disorders in the school context, especially in specific public health situations when wearing a face mask is needed in order to protect the health of the entire community. Objective To estimate the association between the use of protective face masks and risk for voice disorder in university professors.

Methods: This is an observational, analytical, cross-sectional study, conducted in October-November 2022, while using a face mask on campus was mandatory for teachers, students and staff at all times. Initially, all 2251 eligible professors at the Federal University of Santa Catarina (Brazil) received an invitation email to participate in this study. The inclusion criteria made sure that the participants were teaching in-person graduate and/or undergraduate courses in any of the four different University campuses during data collection. Professors who were on any kind of leave of absence at the time of data collection or those who were teaching online for any reason were not included. The research team developed an online questionnaire using the Research Electronic Data Capture (RedCap) platform, that contained questions on sociodemographic data, voice, work characteristics, personal health information, habits and use of masks while teaching. The same questionnaire also included the Screening Index for Voice Disorder (SIVD), a tool that has been developed and validated for screening teachers for the risk of voice disorders[5]. The dependent variable in this study was risk for voice disorder as measured by the SIVD (a score of 5 or more symptoms in a list of 12 indicates risk for voice disorder in teachers). The independent variable was the use of face mask, assessed through the answer to the following question: “do you use a face mask to teach in-person classes?”. Response categories were: “never”, “seldom”, “sometimes”, “always”. The adjustment variables in the study were gender (male, female); age group (29 to 59 or over 60 years of age); weekly hours of class (0-6; 7-9; 10 or more hours); having had COVID-19 (no, yes); having had a positive test for COVID-19 (no, yes); and rest (if participant refers waking up feeling rested). Professors who answered “yes” to wearing face masks (either seldom, sometimes or always) while teaching were asked, subsequently, about what kind of mask and how often they used them (FFP2 or N95, surgical, cloth, or using FFP2/N95 and surgical masks simultaneously). Categorical variables were described in terms of their absolute and relative frequencies, and respective 95% Confidence Intervals (CI95%). Then, the occurrence of the risk of voice disorder was estimated according to the adjustments and independent variables. In order to compare these proportions, either Pearson’s Chi-Square test or Fisher’s Exact test were used. A Logistic Regression Model was applied in order to estimate the association between the use of face masks and risk of voice disorder. For both the bivariate and adjusted analysis the Odds Ratio (OR) was used as the association measurement. The analysis was adjusted for all variables, regardless of the p value. Variables were included for analysis simultaneously using the ‘enter’ method. The Hosmer-Lemeshow (goodness-of-fit test) Test was used in order to test the appropriateness of the final model. For the same purpose, a ROC curve was plotted and the residues were analyzed for symmetry and graphically. The model proved appropriate. The significance level for this study was set at 5% ($p<0.05$) and data analysis was conducted using Stata, version 14. All participants signed an informed consent term in order to take part in this study which was approved by the institutional research ethics committee. All teachers at risk of having voice disorder were assessed and treated by specialist professionals at the University Clinic.

Results: 256 university professors answered the questionnaire and were included in the study. Most of them were women (56.9%), between 29 and 59 years of age (82.1%), with a mean workload of over 10 hours of teaching every week (60.5%). Most of the participants (57.3%) reported having had COVID-19, of whom 51.6% said that they had a positive diagnostic test for COVID-19. Moreover, 85.4% of the professors reported drinking less than 2 liters of water a day, and 62.1% refer waking up feeling rested. Most professors (75.3%) referred always using a face mask while teaching, and according to the results from the SIVD, 36.7% were at risk of having a voice disorder. The most frequently reported vocal symptoms were: strained speech (54.3%), dry throat (46.8%), dry cough (45.3%) and hoarseness (41%), followed closely by phlegm (40.6%). Regarding this instrument's total score, of the teachers who were at risk (5 points or over), 25.01% scored between 5 and 7, and 17.97% scored between 7 and 10. The 12-point maximum score was not achieved by any of the participants. After adjustments to the model, professors who referred that they "sometimes" used a face mask while teaching had 5.65 times more chance of having a voice disorder when compared to those who said they "never" wore a mask while teaching ($p = 0.045$). Those who reported "always" wearing a mask also had more chance of this outcome (OR: 3.55) but without statistical significance ($p = 0.116$). The type of mask used had no significant relation to the risk of voice disorder.

Conclusions: The risk of voice disorder was greater in the participating professors who referred using face masks sometimes while teaching. There was a greater risk of voice disorder in teachers who do not wake up feeling rested. The most frequently reported vocal symptoms were strained speech and dry throat. It is believed that, possibly, either pre-existing or incipient voice disorders may cause teachers to perceive greater difficulty while addressing the class while wearing a mask, which may cause them to either take it off or wear them only in part of the time during class. The authors do not discourage the use of face masks when needed, but we highlight the importance of better understanding the challenges that using a face mask while teaching can cause to these voice professionals who are historically at risk for the development of voice disorders, in order to better prevent and/or treat work-related voice disorders. As the type of mask was not related to the risk of voice symptoms, based on available data on speech intelligibility and protection characteristics, surgical or N95 masks should be preferable for use in the classroom if there is risk of airborne diseases.

References:

1 - Mheidly, Nour et al. Effect of face masks on interpersonal communication during the COVID-19 pandemic. Frontiers in Public Health, p. 898, 2020.

2 - Saunders, Gabrielle H.; Jackson, Iain R.; Visram, Anisa S. Impacts of face coverings on communication: an indirect impact of COVID-19. International Journal of Audiology, v. 60,n. 7, p. 495-506, 2021.

3 - Ribeiro, Vanessa Veis et al. Effect of wearing a face mask on vocal self-perception during a pandemic. Journal of Voice, v. 36, n. 6, p. 878. e1-878. e7, 2022

4 - BOTTALICO, Pasquale et al. Effect of masks on speech intelligibility in auralized classrooms. The Journal of the Acoustical Society of America, v. 148, n. 5, p.2878-2884, 2020

5 - GHIRARDI, Ana Carolina de Assis Moura et al. Screening index for voice disorder (SIVD): development and validation. Journal of Voice, v. 27, n. 2, p. 195-200, 2013

FEATURE SELECTION OF PHONATION PARAMETERS TO DISTINGUISH BETWEEN PRESBYPHONIA AND FES TREATED LARYNGES FROM YOUNG LARYNGES IN AN OVINE MODEL

Bernhard Jakubaß[1,2], Gregor Peters[2], Stefan Kniesburges[2], Marion Semmler[2], Andrijana Kirsch[3], Claus Gerstenberger[3], Markus Gugatschka[3], Michael Döllinger[2]

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

[2] Division of Phoniatrics and Pediatric Audiology at the Department of Otorhinolaryngology, Head and Neck Surgery, University Hospital Erlangen, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

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

***Keywords*: Ex vivo; Machine learning; Presbyphonia; Sheep**

Abstract:

Objectives / Introduction: With age, the voice quality decreases owing the age-related atrophy of the laryngeal muscles. This is called presbyphonia. In previous studies [1, 2] the beneficial impact of the Functional Electrical Stimulation (FES) treatment could be shown. In this study, the best parameter set to distinguish between aged (presbyphonia), young, and stimulated (treated) voices is evaluated in an ex vivo ovine model.

Methods: The stimulation electrodes were placed at the recurrent laryngeal nerve in 12 old sheep (10 years) to stimulate the thyroarytenoid muscle (TAM). Stimulation was applied to 6 sheep for nine weeks. For the phonation experiments, the harvested larynges were fixated in a mechanical setup and the vocal folds (VF) were postured in phonation position. Each larynx performed 64 measurement runs with different mass-flow and pre-stress settings. For each case, the VFs vibration was recorded with a highspeed camera. Simultaneously, the subglottal pressure and the sound were measured. From these signals, 28 parameters were computed using the Glottis Analysis Tools (GAT) [3] software package [1]. Additionally, the volume of the TAM and the length of the vocal folds was segmented from microCT scans [2]. The data of the aged non-stimulated sheep were augmented with data of similar aged (9 years) unstimulated sheep from a pre-study [4]. Additionally, the data from another pre-study of unstimulated but young (younger than one year) sheep [5] was added as a third group. The experiments, data collection, and data analysis of both pre-studies [4, 5] was done in the same way as the new data [1, 2].

To find the best parameter set to distinguish between the three groups, a feature selection using SVM classifiers was performed, utilizing MATLAB. In total, 1006 datasets (486 aged, 377 stimulated, 143 young) with 30 parameters were used. The performance of each classifier was evaluated using five-fold cross validation. Each parameter was left out once and the best parameter set was used in the next iteration. The final classifier was then tested on a separate test set.

Results: After 24 iterations with seven parameters left, the loss of the classifier increased drastically, see Figure 1. These seven parameters are the mass-flow (Q), fundamental frequency ($f_{0,audio}$), Amplitude Periodicity (AP), Signal-toNoise-Ratio (SNR_{audio}), Cepstral Peak Prominence (CPP_{audio}), jitter ($Jitt\%_{Psub}$), and the total TAM volume, see Figure 2. All named parameters show statistically significant differences between the three groups (Kurskal-Wallis test, $p < 0.001$). The overall accuracy of the classifier on the test set is 86.6 %. For the young group the accuracy value even reaches 90.3 %.

Conclusions: As Q and $Jitt\%_{Psub}$ representing the aerdynamics and SNR_{audio} and CPP_{audio} representing the acoustic signal, it may be concluded that the glottal dynamic signal contributes less to the separation of the three groups, as AP is the only parameter in the remaining parameter set. The $f_{0,audio}$ cannot be counted as the acoustic signal as it represents the (identical) fundamental frequency of all three signals. With the limited amount of data in mind also the overall accuracy of the classifier seems reasonable.

Acknowledgements:

The project was supported by the Deutsche Forschungsgemeinschaft (DFG, Germany) Grant No. DO1247/12-1 and the Fonds zur Förderung der wissenschaftlichen Forschung (FWF, Austria) Grant No. I3997.

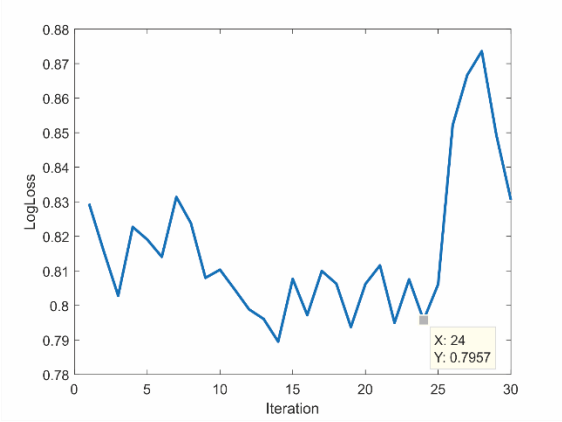


Figure 1: LogLoss of SVM classifier over iterations.

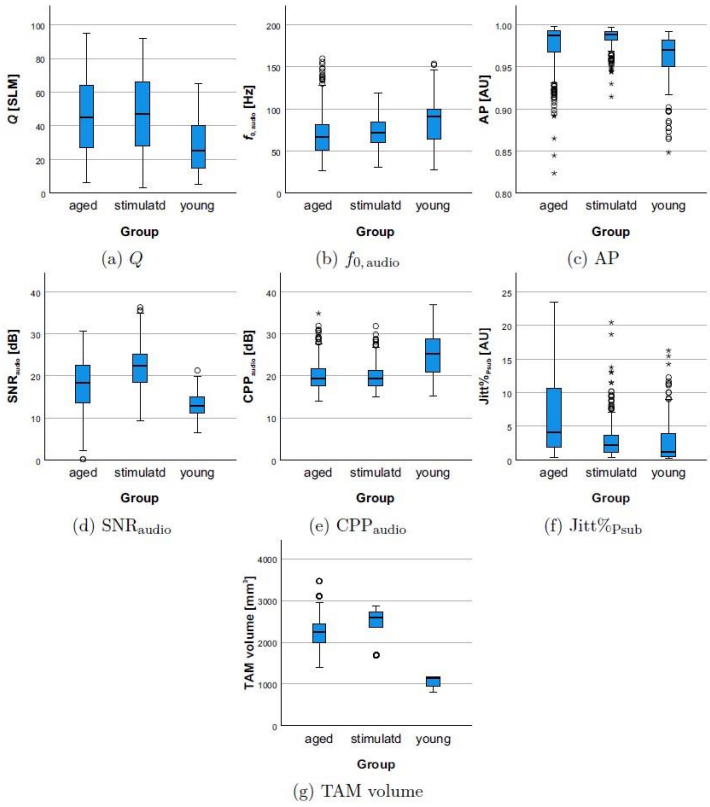


Figure 2: Boxplots of the final parameter set.

References:

[1] B. Jakubaß, G. Peters, S. Kniesburges, M. Semmler, A. Kirsch, C. Gerstenberger, M. Gugatschka, M. Döllinger. Effect of functional electric stimulation on phonation in an ex vivo aged ovine model. Journal of the Acoustic Society of America, 153(5):2803–2817, 2023.

[2] A. Kirsch, C. Gerstenberger, B. Jakubaß, M. Tschernitz, J. Perkins, A. Groselj-Strele, H. Lanmüller, J. C. Jarvis, S. Kniesburges, M. Döllinger, M. Gugatschka. Bilateral functional electrical stimulation for the treatment of presbyphonia in a sheep model. The Laryngoscope, 134(2):848-854, 2023.

[3] A. M. Kist, P. Gómez, D. Dubrovskiy, P. Schlegel, M. Kunduk, M. Echternach, R. Patel, M. Semmler, C. Bohr, S. Dürr, A. Schützenberger, and M. Döllinger. A deep learning enhanced novel software tool for laryngeal dynamics analysis. Journal of Speech, Language, and Hearing Research, 64(6):1889–1903, 2021.

[4] C. Gerstenberger, M. Döllinger, S. Kniesburges, V. Bubalo, M. Karbiener, H. Schlager, H. Sadeghi, O. Wendler, and M. Gugatschka. Phonation analysis combined with 3d reconstruction of the thyroarytenoid muscle in aged ovine ex vivo larynx models. Journal of voice: official journal of the Voice Foundation, 32(5):517–524, 2018.

[5] M. Döllinger, O. Wendler, C. Gerstenberger, T. Grossmann, M. Semmler, H. Sadeghi, and M. Gugatschka. Juvenile ovine ex vivo larynges: Phonatory, histologic, and micro ct based anatomic analyses. BioMed Research International, 2019, 2019.

A MORE COMPLETE VIEW INTO THE VOCAL FOLD TISSUE DURING EX-VIVO PHONATION EXPERIMENTS BY USING A COMBINATION OF PIPETTE ASPIRATION AND ULTRASOUND ELASTOGRAPHY

Florian Scheible[1], Raphael Lamprecht[1], Reinhard Veltrup[2], Jann-Ole Henningson[3], Marion Semmler[2], Alexander Sutor[1]

[1]Institute of Measurement and Sensor Technology, UMIT TIROL – Private University for Health Sciences and Health Technology, Hall in Tirol, Austria

[2]Division of Phoniatrics and Pediatric Audiology, Department of Otorhinolaryngology, Head- and Neck surgery, University Hospital Erlangen, Friedrich-Alexander-University Erlangen-Nürnberg, Erlangen, Germany

[3]Chair of Visual Computing, Friedrich-Alexander-University Erlangen-Nürnberg, Erlangen, Germany

***Keywords*: Ex-vivo experiment; Tissue Phonation; Pipette Aspiration Technique; US Elastography**

Abstract:

Objectives / Introduction:

In order to control the phonation during speech the vocal folds undergo continues muscular manipulation [1]. These manipulations do not only alter the vocal fold constellation but also the biomechanical properties of the tissue. The vocal folds are a layered structure, which can be simplified into a cover and body part according to Hirano [2]. To determine the change in both layers, it is necessary to measure in different depths of the sample which requires a combination of measurement techniques. Whereby most of the material testing make a dissection of the sample necessary. In contrast, the pipette aspiration technique and ultrasound elastography offer an approach which is noninvasive and through its combination able to produce a more complete view inside of the vocal folds tissue. Thus, it is possible to study the effect of different manipulations on the tissue, which provides new insights into understanding the process of speech production.

Methods:

In ex-vivo experiments on ten larynges, seven different manipulations were applied as adduction and elongation, and the larynges were brought to phonation by a controlled airflow through the remaining trachea. The material properties of the tissue were determined for each manipulation. With the pipette aspiration technique for the near surface area and with the US elastography of the vocal musculus, both measurement techniques are well established in voice science and can be seen in Figure 1 [3,4]. Further the movement of the vocal folds were recorded with a highspeed camera, so variations in the phonation can be identified [5]. In addition, the subglottal pressure as well as the produced sound is recorded.

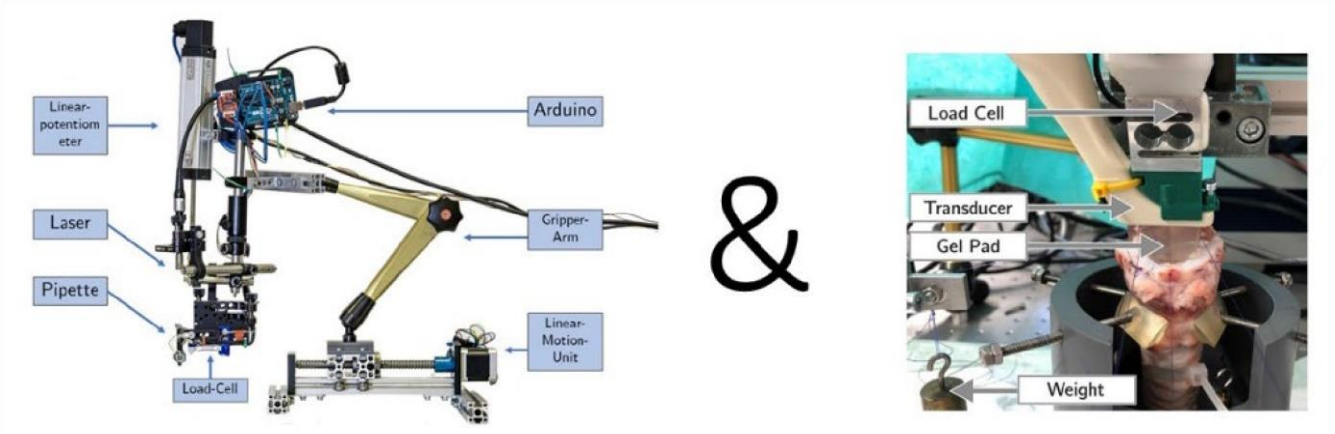


Figure 1: Both measurement devices are shown above. The pipette aspiration setup with its named components on the left as well as the USelastography setup on the right.

Results:

The applied manipulations do affect the elasticity of the tissue. Regardless of its region the manipulation led to higher elasticity values. Furthermore, it can be seen in Figure 2, that the tissue of the body, which is measured with the USelastography is in general stiffer as the cover is more effected by the manipulations as the cover, measured with the pipette used as an indenter for static elasticity measurements. The pipette aspiration data is also shown in the plot but can not be compared directly because the dynamic measurement differs to the static ones made by the two other techniques.

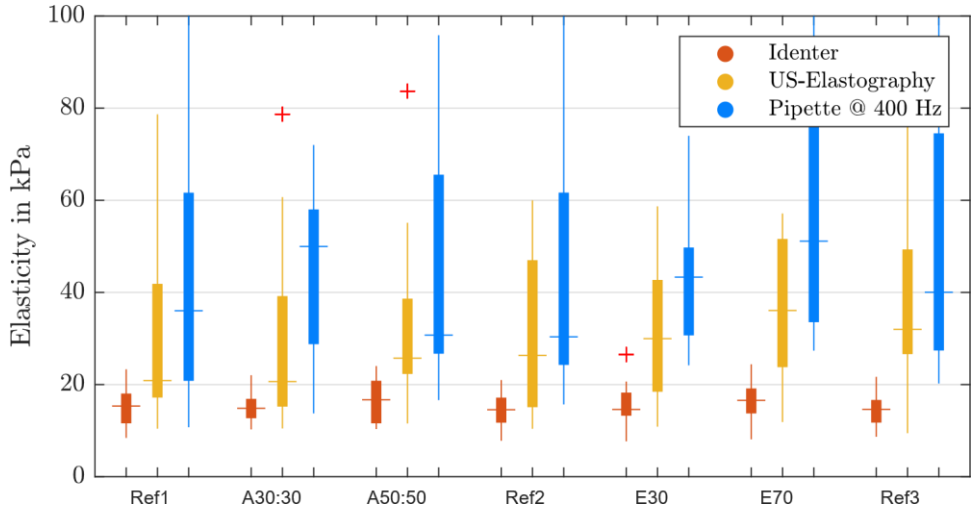


Figure 2:The static elasticity values for both vocal folds sides determined with the US-elastography and the pipette used as an indenter for static measurements and in its purpose for dynamic measurements are plotted to compare them. The measurement series starts with a reference measurement followed by symmetric adductions with 30g and 50g. After a renewed reference measurement, the larynx is elongated with 30 g and 70g. Finaly a last reference measurement is made.

Conclusions:

It was successfully possible to integrate both measurement techniques into ex-vivo phonation experiments and receive material properties of the vocal fold on different locations.

The data of all measurement techniques show response to the manipulations, where the highest elasticity rise can be seen in the vocalis muscle.

The results of this combined measurements give a novel insight in to the understanding of the process of phonation.

Acknowledgements:

This work was supported by the Austrian Science Fund Grant No. I 3806-B28 and by the German research foundation (DFG) No. DO1247/9-1. All authors declare that they have no conflicts of interest.

References:

[1] Sundberg, J. The Acoustics of the Singing Voice, Scientific American offprints, Scientific American, 1977

[2] Hirano, M. Morphological structure of the vocal cord as a vibrator and its variations. Folia phoniatrica 1974, 26, 89–94. 158 doi: /10.1159/000263771

[3] Scheible, F.; Lamprecht, R.; Schaan, C.; Veltrup, R.; Henningson, J.O.; Semmler, M.; Sutor, A. Behind the Complex Interplay of Phonation: Investigating Elasticity of Vocal Folds With Pipette Aspiration Technique During Ex Vivo Phonation Experiments. Journal of Voice 2023. doi:10.1016/j.jvoice.2023.03.001

[4] Lamprecht, R.; Scheible, F.; Veltrup, R.; Schaan, C.; Semmler, M.; Henningson, J.O.; Sutor, A. Quasi-static ultrasound elastography of ex-vivo porcine vocal folds during passive elongation and adduction. Journal of Voice 2022 doi:10.1016/j.jvoice.2022.11.033

[5] Semmler, M.; Berry, D. A.; Schützenberger, A; Döllinger, M. Fluid-structure-acoustic interactions in an ex vivo porcine phonation model, The Journal of the Acoustical Society of America 149 (2021) 1657–1673. doi:10.1121/10.0003602.

Optimizing Frequency Response in Ambulatory Voice Monitoring: Aligning Laboratory Accelerometer Data for Enhanced Voice Disorder Assessment

Víctor M. Espinoza[1], Juan P. Cortés[2], Rocio Ortega[2,3], Christian Castro[1,2], Matías Zañartu[2,3]

[1] Department of Sound, Universidad de Chile, City, Santiago, Chile

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

[3] Department of Electronic Engineering, Universidad Técnica Federico Santa María, Valparaíso, Chile

***Keywords*: Ambulatory Voice Monitoring; Equalization; Neck-Surface Acceleration**

Abstract:

Objectives / Introduction:

In recent years, there has been an active area of research on ambulatory devices for voice monitoring, specifically to study vocal behavior and/or vocal pathologies [1]. These have been used in commercial and research settings [2], where the main component is a miniature accelerometer that it attaches to the neck-surface of the user. When using ambulatory voice monitoring devices (a prototype shown in Figure 1A to the right), a variation in frequency and sensitivity response of the neck skin acceleration (NSA) is observed in relation to a light-weight accelerometer used in laboratory (in-lab) settings, shown at the left in Figure 1 A. Since current ambulatory devices might have material properties covering the accelerometer sensor, the frequency response of these devices will also differ. In order to fairly compare data extracted from different devices, we propose a method to match the frequency response of ambulatory devices to the frequency response of an in-lab accelerometer setup. Therefore, any measure or feature extracted from the in-lab accelerometer that is frequency-dependent could be compared with the same measures/features extracted from a given device. The case of inverse filtering the accelerometer signal to obtain an estimation of the glottal flow is an example where the frequency response of the sensors (in-lab and ambulatory device) should be as similar as possible to avoid differences and bias from the extracted voice features. Thus, the aim of this study is to correct for the frequency response difference between laboratory and in-field neck surface acceleration for ambulatory monitoring devices.

Methods:

A proof-of-concept experiment was conducted to examine two configurations of the Knowles BU-27135 accelerometer, i.e., an in-lab and an ambulatory voice monitoring (AVM) prototype. The accelerometers were mounted on a 3" full-range speaker using a double-sided tape (Model 2181, 3M, Maplewood, MN), a setup that produces a piston-like vibration at frequencies up to 1 kHz. A high-bandwidth test signal was then applied to the speaker, which produced a surface acceleration over both accelerometers. Following the methods proposed in [3], a sweep test signal allows for improving the signal-to-noise ratio and for reducing distortion measurements. To obtain the frequency response and eliminate the non-linear components of the speaker, the measured signals were deconvolved via inverse filtering using the Discrete Fourier Transform (DFT), as proposed in [3].

Figure 1 (A): An accelerometer packed in a light-weight silicone (left side) and an AVM prototype (right side). Both sensors have the same accelerometer brand and model. The AVM includes a portable audio digital interface.(B): Setup of both accelerometers attached to the speaker with a double-sided 3M tape.

Results: Figure 2 (upper panel) shows the magnitude response of both sensors for this initial experiment. As observed, between 200 and 800 Hz, both sensors have a nearly flat frequency response. However, below 200 Hz, the AVM needs a gain boost of up to 10 dB to match the in-lab accelerometer frequency response. A similar case occurs at frequencies higher than 800 Hz, with correction gains up to 6 dB approximately. In order to compensate for these gains,

we can use the coefficients of the impulse response derived from the DFT-based data as a FIR filter (in our experiment, 512 coefficients), or we can estimate a stable auto-regressive moving average (ARMA) model to implement an IIR filter, which reduces the computational cost in these devices. The resulting compensation ARMA filter for this experiment is shown in Figure 2 (lower panel) with an estimated order of poles (p) and zeros (q) of (p,q) = (6,6).

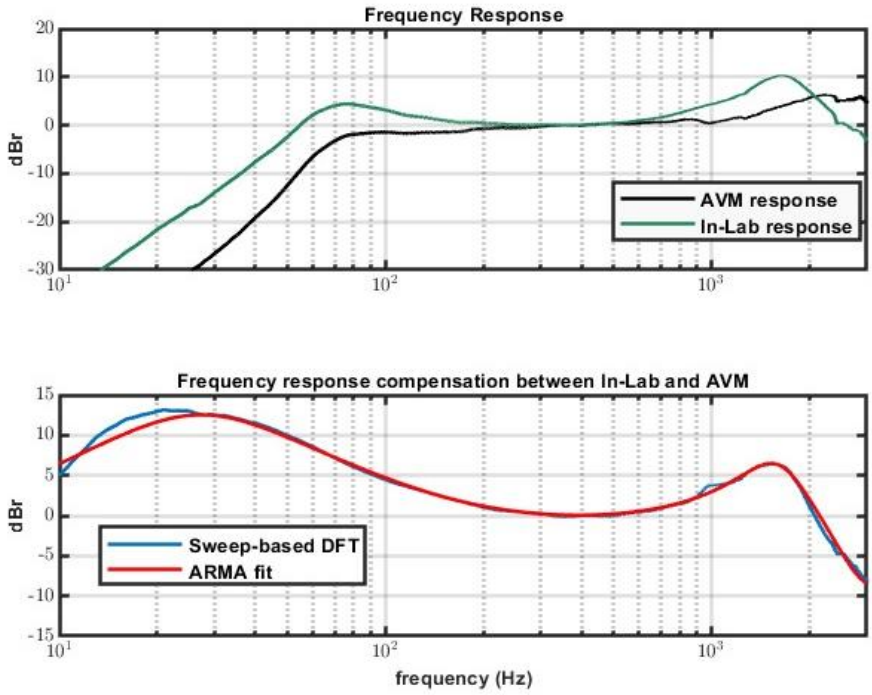


Figure 2: Differences between similar accelerometers mounted in different packages (See Figure 1). At top, Note the magnitude frequency response changes at low and high frequencies, which means the AVM needs to compensate for these differences to improve the accuracy between inlab v. infield voice recordings. At bottom, the frequency response to compensate in the AVDM, including an ARMA fit (p,q)=(6,6) from the sweep-based DFT data.

Conclusions: This preliminary experiment emphasizes the importance of specific calibration or frequency compensation for the AVM and similar devices that include an accelerometer. By compensating for these responses, direct comparisons between results obtained from accelerometers used in laboratory experiments and those used in the field (i.e., ambulatory) are possible. This experiment has the potential to aid in the design of repeatable methods to provide researchers with a reliable and simple method for accurately comparing results obtained from laboratory conditions with those from ambulatory voice monitoring devices.

Acknowledgments:

This study was supported by ANID grants FONDECYT 1241142 and BASAL FB0008. The authors acknowledge Lanek SpA for kindly providing the AVM prototype used in this study.

Dr. Matías Zañartu has a financial interest in Lanek SPA, a company focused on developing and commercializing biomedical devices. Dr. Zañartu's interests were reviewed and are managed by Universidad Técnica Federico Santa María in accordance with its conflict-of-interest policies.

References:

[1] D. D. Mehta et al., 'Using ambulatory voice monitoring to investigate common voice disorders: research update', Front. Bioeng. Biotechnol. 3:155. doi: 10. 3389/fbioe. 2015. 00155, 2015.

[2] J. H. Van Stan, J. Gustafsson, E. Schalling, and R. E. Hillman, 'Direct Comparison of Three Commercially Available Devices for Voice Ambulatory Monitoring and Biofeedback,' Perspectives on Voice and Voice Disorders, vol. 24, p. 80, 2014.

[3] A. Novak, P. Lotton, and L. Simon. Synchronized Swept-Sine: Theory, Application, and Implementation. Journal of the Audio Engineering Society, 63(10):786–798, November 2015. Publisher: Audio Engineering Society.

ACOUSTIC CHARACTERISTICS OF THE SINGING VOICE AND VOCAL TRACT CHANGES DURING SINGING INSTRUCTION

Jun Takahashi[1], Itsuki Shishime[2], Natsuki Toda[2], Hironori Takemoto[2]

[1] Osaka University of Arts, Osaka, Japan

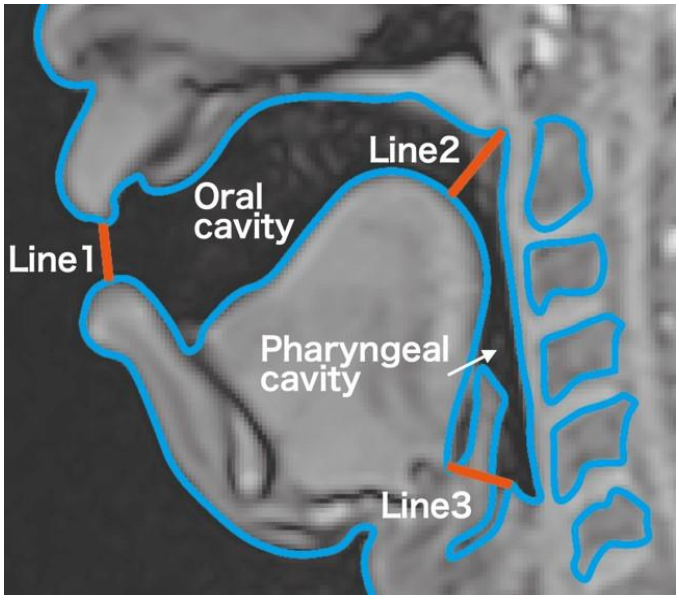
[2] Chiba Institute of Technology, Chiba, Japan

***Keywords*: Singing Voice; Vocal Tract; Real-Time MRI; Singer's Formant**

Abstract:

Objectives / Introduction: In singing instruction, internal body movements during singing cannot be observed directly. Therefore, teachers judge the state of the vocal organs based on the singing voice and use words to describe internal body movements (such as "open your throat"). Recently, real-time magnetic resonance imaging (rtMRI) has made it possible to capture moving images. This approach can be used to observe the movements of the vocal organs during singing [1]. In this study, we evaluated changes in acoustic characteristics and body movements during singing after one year of singing instruction.

Methods: The vocal participant was a student with no previous singing training. Singing instructions were provided by a professional vocalist who taught at the music academy. A total of 24 lessons were conducted over the course of one year. The subject underwent rtMRI and voice recording before beginning training. Voice recordings were then made monthly. Finally, rtMRI and voice recording were performed at the end of the study period. The singing task involved the vowel /a/ at any pitch. Images of the vocal tract were obtained using an MRI system (Siemens MAGNETOM Prisma Fit 3T). The subject sang in the supine position. Movies were captured for 50 s at a slice thickness of 10 mm and a frame rate of 10 fps. The pixel size was 1 × 1 mm. The contours of the vocal tract were extracted as follows: line 1 was the shortest line between the upper and lower lips, line 2 was the shortest line between the contact point of the soft palate and the posterior wall of the pharynx to the tongue, and line 3 connected top of the arytenoid region to the bottom of the vallecula. The vocal tract bound by line segments 1 and 2 was determined as the width of the oral cavity, and that Fig. 1 Methods for extracting vocal tract contours bound by line segments 2 and 3 was determined as the width of the pharyngeal cavity in pixels (Fig. 1). The Singer's formant (SF), a common acoustic characteristic, is observed at approximately 3 kHz in the spectrum of the singing voice; the more pronounced it is, the higher it is evaluated by the listener. In this experiment, we used STRAIGHT [2] to analyze singing voices. In the STRAIGHT analysis, a singing voice is decomposed into a fundamental frequency and a smoothed spectrum that represents the characteristics of the vocal tract. The sum of the power in the 0–12 kHz band and the sum of the power in the SF band (2–4 kHz) were determined. The value of SF power divided by the total power was defined as "SF occupancy." Thus, if the total power was in the 2–4 kHz band, the SF occupancy was 1 (maximum). The vocal participant was evaluated by 18 professional singers and instrumentalists (three sopranos, three tenors, six baritones, five pianists, and one violinist). The singing voice was rated on a 5-point Likert scale (Good, 5; Bad, 1). Mean ratings of the singing voice before and after training were compared using a t-test.



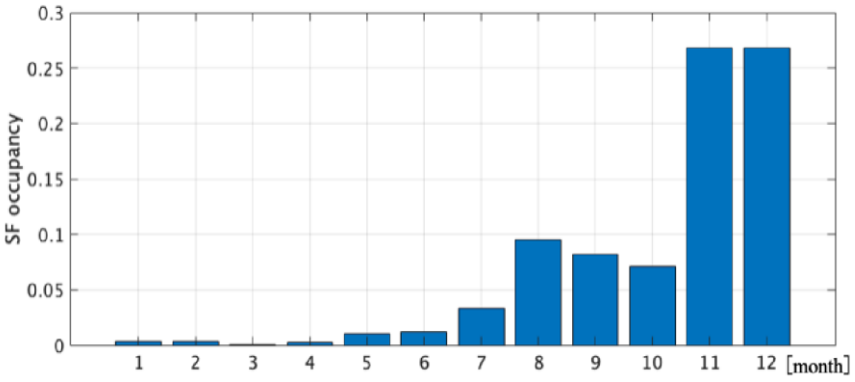


Fig. 2 SF occupancy rate by month

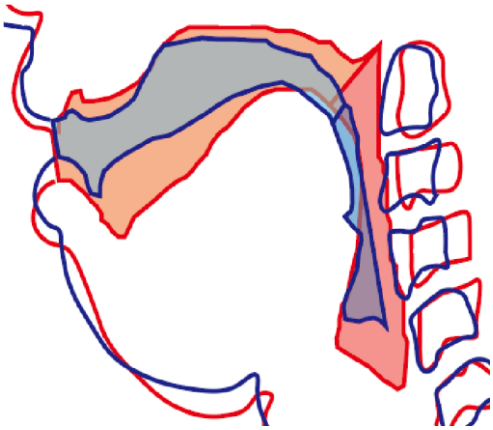


Fig. 3 Vocal tract shape before and after training

Results: The mean ratings of the singing voice were 2.22 before training and 3.44 after training, and their difference was significant ($t(17) = -3.42$, $p = 0.0033$), suggesting that singing voices improved after training. As shown in Fig. 2, SF occupancy rates of the singing voice were 0.0033 before training and 0.2680 after training, indicating an increase after training. These results explain the increase in singing voice ratings after training. It can be observed that the SF occupancy rate increased gradually after the training. Fig. 3 shows the traced and layered vocal tract shapes of the vowel /a/ before and after training. The oral and pharyngeal cavities were larger after training (red) than before training (blue). Before training, the oral cavity was 1298 pixels and the pharyngeal cavity was 383 pixels. After training, the sizes were 2183 and 985 pixels, respectively. These results indicate that one year of singing training expanded the oral and pharyngeal cavities. The expansion of the oral cavity could be achieved mainly by widely opening the mouth, while the expansion of the pharyngeal cavity mainly by lowering the larynx. The shape of the vocal tract, particularly the large cross-sectional area ratio of the larynx to pharynx (which is necessary for SF synthesis [3]), is related to the acoustic characteristics of the singing voice. Therefore, the expansion of the pharyngeal cavity after training increased the SF occupancy of the singing voice.

Conclusions: In this study, we examined how the acoustic characteristics of a student's singing voice and vocal tract changed after one year of singing training. Evaluation by 18 professional singers and instrumentalists rated the singing voice higher after training than before. This suggests that one year of singing training caused a change in the singing voice of the subject, resulting in a higher evaluation. A singing voice analysis also showed that the SF occupancy of the singing voice was higher after the training than before, providing insight into the improvement in singing voice after training. Furthermore, the oral and pharyngeal cavities were larger after training than before training. The shape of the vocal tract is related to the acoustic characteristics of the singing voice, and the expanded pharyngeal cavity after training increases the SF occupancy of the singing voice.

Acknowledgements:
This work was supported by JSPS KAKENHI (Grant Numbers 22K13773 and 23K11172). We also thank all of the participants in the experiments.

References:

[1] Takemoto, H., Goto, T., Hagihara, Y., Hamanaka, S., Kitamura, T., Nota, Y., & Maekawa, K. (2019). *Speech organ contour extraction using real-time MRI and machine learning method*. Proceedings of Interspeech 2019, 904-908. Graz, Austria.
[2] Kawahara, H., Masuda-Katsuse, I., & de Cheveigné, A. (1999). *Restructuring speech representations using a pitchadaptive time–frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds.* Speech Communication, 27, 187-207.
[3] Sundberg, J. (1987). *The Science of the Singing Voice*. Northern Illinois University Press.

Consideration of the vocal tract acoustics of different voice qualities with regard to voice efficiency

Fiona Stritt[1,2], Stefanie Rummel[3], Johannes Fischer[2,4], Michael Bock[2,4], Bernhard Richter[1,2], Matthias Echternach[5], Mario Fleischer[6*], Louisa Traser[1,2*]

[1]Institute of 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
[6]Charité – Universitätsmedizin Berlin, Department of Audiology and Phoniatrics, Germany

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

Abstract:

Objectives / Introduction:

In the western classical style of singing, aesthetic aspects are very closely linked to economic aspects of voice production, since the singer must be audible through an orchestra without technical amplification. In this context, a voice production is defined as efficient, when its mechanism achieves a high vocal sound pressure level compared to the energy spent and the bio-mechanical properties employed by the singer. Effectiveness is influenced by breathing strategy, voice source (e.g., spectral slope), and vocal tract (VT) acoustics (e.g., amplification of psychoacoustically relevant frequency sections in the spectrum) (1) as well as their interactions (2). In contemporary commercial music, a considerably wide range of singing styles is used. Thanks to electronic amplification and modern recording technology, it is also possible to apply very soft voice production types in noisy environments. However, if singers use a less efficient strategy to produce a high vocal output the risk of vocal overload increases. To quantify voice efficiency, our group recently proposed a measure for acoustic sound intensity within the human glottis depending on different vocal tract configurations and vocal fold vibration (3). This measure was applied in an initial single-subject study aiming to quantify voice efficiency in 6 voice qualities (Speech, Falsetto, Sob, Oral Twang, Opera, and Belting) as defined by Estill Voice Training® (4). In this continuation we aim to analyze the transferability of the proposed measure to different individuals, beginning with the conjunction of VT configuration and its acoustic properties.

Methods:

Four professional singers (two females – subjects A&B and two males – subjects C&D) with a high qualification level in Estill Voice Training® were included in this study and asked to perform sustained phonation on vowel [a:] in 6 voice qualities on G#4 (415Hz) for females and G#5 (207Hz) for males during 3D MRI recordings of the vocal tract as described in (5). Segmentation of MRI data, implementation of teeth models, preparation of the model for simulation as well as calculation of acoustic properties in transfer functions by finite-element-modeling was performed as described in (6). The volume and partial volume of defined VT segments were analyzed based on anatomic landmarks (7). Based on this the ratio between the mean oropharyngeal and hypopharyngeal volume (OP/HP ratio) was calculated. Vocal tract length was determined according to (8).

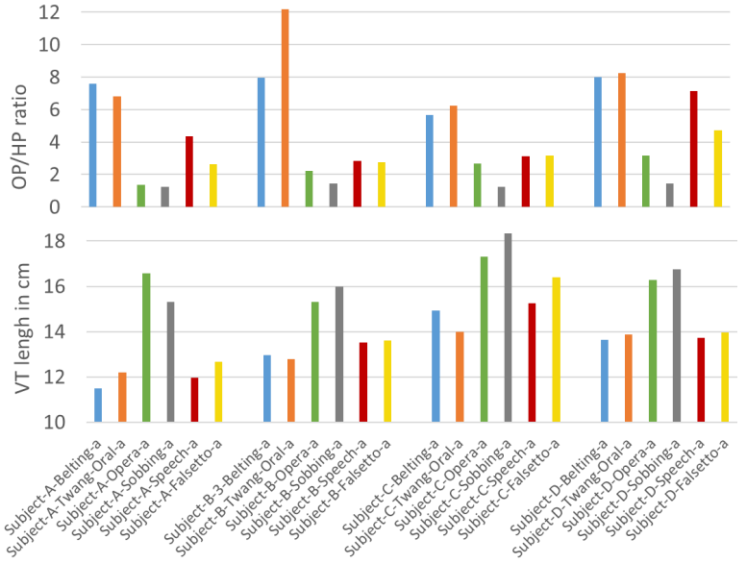


Figure 1: *Ratio of oro- vs. hypopharyngeal Volume (first row) and vocal tract length (second row) for all subjects and voice qualities.*

Results:

Singers presented characteristic VT configuration for the respective voice quality with a high level of inter-individual agreement in volumetric analysis (see Figure 1): Belting and Twang had a similar configuration and were characterized by a megaphone shape (hypopharyngeal constriction with trumped-shaped widening with a high OP/HP ratio and low VT length). Opera and Sobbing in contrast were characterized by an anti-megaphone configuration including a hypopharyngeal widening with a consequently reduced OP/HP ratio and a low laryngeal position), most pronounced for Sobbing. Falsetto and Speech were intermediate in both, length and configuration, however, exhibited the highest individual variance.

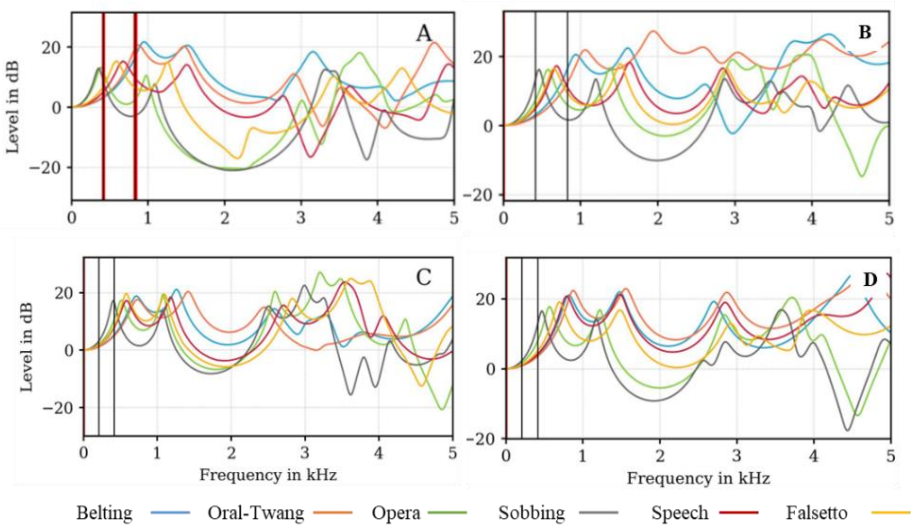


Figure 2: *Vocal tract transfer functions for all subjects and voice qualities. Horizontal lines representing the fundamental frequency and the first partial*

This grouping was also reflected in the VT acoustics with a strong association of voice quality with the position of the first two resonances ($f_{R1/2}$) and thus vowel color (see Figure 2). An efficiency gain can be discussed for the association of f_{R1} with f_0 for Opera and Sobbing and of f_{R1} with $2f_0$ for Belting and Twang for the female subjects (A&B).

An enhancement of harmonic energy in the psycho-acoustically relevant region of 2-4 kHz was pronounced in Opera and Belting (for all subjects) as well as Twang (for subjects B&D). Sobbing was in contrast characterized by the occurrence of anti-resonances in this region. In Falsetto/Speech, inter-individual differences in the acoustic evaluation became apparent showing no specific strategy for enhancing acoustic energy.

Conclusions:

The analysed six voice qualities differed concerning their VT configuration and related acoustic properties. The four singers showed consistent vocal tract configurations and acoustic strategies, especially for Belting, Twang, Opera, and Sobbing. Falsetto and Speech exhibited stronger individual variations suggesting that these qualities are less defined by their acoustic properties. In accordance with (3) the transfer functions indicated that Belting and Opera represent different but particularly efficient strategies as both enhance acoustic energy in psycho-acoustically relevant regions. Sobbing was characterized by an energy loss through the occurrence of antiresonances possibly caused by the piriform sinus and the vallecula. The significance of this study is limited due to the small number of subjects. Still, the uniformity of the vocal tract configurations and their acoustic properties in the reproduction of six voice qualities in four different individuals indicates that it is possible through training to apply defined resonance strategies with specific properties in terms of their efficiency. In the next step, additional consideration of the VT efficiency in relation to the voice source based on (3) is planned.

Acknowledgements:

Louisa Traser (TR1491/4-1), Johannes Fischer (FI 2803/1-1), Matthias Echternach (grant Ec409/1-4) are supported by the Deutsche Forschungsgemeinschaft (DFG).

References:

1. Sundberg J. The science of the singing voice. Northern Illinois University Press; 1987.
2. Titze IR, Story BH. Acoustic interactions of the voice source with the lower vocal tract. 2014;101(4):2234–43.
3. Fleischer M, Rummel S, Stritt F, Fischer J, Bock M, Echternach M, et al. Voice efficiency for different voice qualities combining experimentally derived sound signals and numerical modeling of the vocal tract. Front Physiol. 2022;13.
4. Steinhauer K, Klimek M, Estill J. The Estill Voice Model: Theory and Translation. 2017.
5. Traser L, Birkholz P, Flügge TV, Kamberger R, Burdumy M, Richter B, et al. Relevance of the Implementation of Teeth in Three-Dimensional Vocal Tract Models. J Speech Lang Hear Res. 2017;60(9):2379–93.
6. Birkholz P, Kürbis S, Stone S, Häsner P, Blandin R, Fleischer M. Printable 3D vocal tract shapes from MRI data and their acoustic and aerodynamic properties. Sci Data. 2020;7(1):1–16.
7. Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, et al. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. Neuroimage. 2006 Jul;31(3):1116–28.
8. Echternach M, Birkholz P, Traser L, Flügge T V., Kamberger R, Burk F, et al. Articulation and vocal tract acoustics at soprano subject's high fundamental frequencies. J Acoust Soc Am. 2015;137(5):2586–95.

METAGENOMIC WHOLE GENOME SHOTGUN ANALYSIS OF THE AIRWAY MICROBIOME IN LARYNGOTRACHEAL STENOSIS

Nour Awad, MD[1], Peter J. Larson, MD, PhD[1], Cheick A. Sissoko, MSc[1], Laurel L. Ball, MD[1], Gregory R. Dion, MD, MS, FACS[1]

[1] University of Cincinnati College of Medicine, Department of Otolaryngology – Head and Neck Surgery, Cincinnati, OH, USA

***Keywords*: Laryngotracheal stenosis, larynx, microbiome, metagenomic whole genome sequencing**

Abstract:

Objectives / Introduction: The airway microbiome has been implicated in the pathogenesis of laryngotracheal stenosis (LTS).[1-4] However, these studies have been limited by the sub-genus-level resolution of amplicon-based 16S rRNA sequencing. Metagenomic whole genome shotgun sequencing (mWGS), by contrast, allows for strain-level taxonomic and functional genomic analysis, providing insight into the specific organisms and pathways involved. This study aims to explore the role and function of microbiome in LTS using mWGS.

Methods: mWGS was conducted on 12 intraoperative swab samples from 4 patients with LTS and 4 control patients. Patients also underwent chart review and completed questionnaires to document their medical history, antibiotic use, zip code, diet, and current medications. DNA was extracted and sequenced on an Illumina NextSeq 2000. Using Biobakery workflows, reads were quality controlled through KneadData and taxonomic profiles were generated through MetaPhlAn4.

Results: LTS samples exhibited decreased taxonomic diversity and were dominated by species with previously described roles in other chronic inflammatory processes such as Streptococcus salvarius, Streptococcus mitis, Anaerococcus vaginalis. Our results demonstrate species- and strain-level microbiome differences in patients with LTS and shed light on potential microbial players in the chronic inflammatory processes of airway stenosis (Figure 1). These data illustrate the additional insight mWGS can provide and enable downstream strain-level and functional analysis to characterize the pathogenesis of LTS.

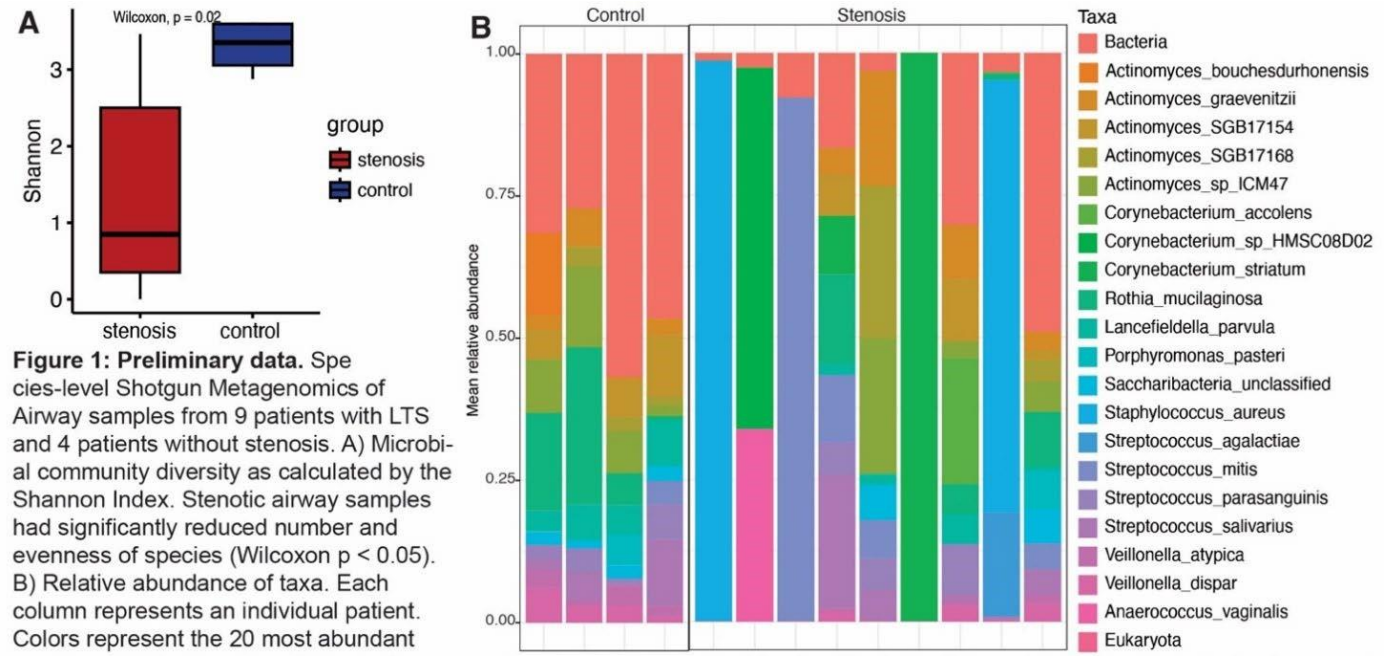


Figure 1: Preliminary data. Species-level Shotgun Metagenomics of Airway samples from 9 patients with LTS and 4 patients without stenosis. A) Microbial community diversity as calculated by the Shannon Index. Stenotic airway samples had significantly reduced number and evenness of species (Wilcoxon $p < 0.05$). B) Relative abundance of taxa. Each column represents an individual patient. Colors represent the 20 most abundant species, with the less abundant species grouped under their respective kingdon (Bacteria, Eukaryota). LTS patients had a decreased proportion of low-abundance taxa and a predominance of species such as *Staphyloccocus aureus*, *Streptococcus mitis, Streptococcus parasanguinis*, and Actinomyces species. The Strain level biology of these organisms is a rich and growing field of research. We will be able to map our dataset to existing strain databases and apply current understanding of these organisms in the context of LTS.

Conclusions: Analyzing microbiota variations in LTS could reveal etiology-specific biomarkers, comorbidity links, and therapeutic targets.

Acknowledgements:

The Authors would like to acknowledge the expertise, advise, and support from Dr. Xiang Zhang, PhD, Core Director of the Genomics, Epigoenomics and Sequencing Core and Associate Professor in the Department of Environmental Health at the University of Cincinnati.

References:

1. At H, Ss T, C C, et al. Laryngotracheal microbiota in adult laryngotracheal stenosis. mSphere. 2019;4(3). https://pubmed-ncbi-nlm-nih-gov.uc.idm.oclc.org/31043518/. Accessed Dec 24, 2023. doi: 10.1128/mSphereDirect.00211-19.
2. Fan Z, Zhang L, Wei L, Huang X, Yang M, Xing X. Tracheal microbiome and metabolome profiling in iatrogenic subglottic tracheal stenosis. BMC Pulmonary Medicine. 2023;23. https://www-ncbi-nlm-nihgov.uc.idm.oclc.org/pmc/articles/PMC10523634/. Accessed Dec 24, 2023. doi: 10.1186/s12890-023-02654-7.
3. Mr A, S AAG, Pm G, In J, R G. Drug-eluting endotracheal tubes for preventing bacterial inflammation in subglottic stenosis. The Laryngoscope. 2022;132(7). https://pubmed-ncbi-nlm-nih-gov.uc.idm.oclc.org/34319583/. Accessed Dec 24, 2023. doi: 10.1002/lary.29769.
4. Mr A, A M, Rm F, et al. Amelioration of subglottic stenosis by antimicrobial peptide eluting endotracheal tubes. Cellular and molecular bioengineering. 2023;16(4). https://pubmed-ncbi-nlm-nih-gov.uc.idm.oclc.org/37811005/. Accessed Dec 24, 2023. doi: 10.1007/s12195-023-00769-9.

VOCAL TRANSIENT ANALYSIS USING PIECEWISE LINEAR APPROXIMATION

Takeshi Ikuma, PhD[1,2], Andrew J. McWhorter, MD[1,2], Robin Samlan[5], Tobias Schraut[4], Michael Doellinger[4], and Melda Kunduk, PhD[1,2,3]

[1]Dept. of Otolaryngology-Head and Neck Surgery, LSU Health Sciences Center, New Orleans, Louisiana, U.S.A.
[2]Voice Center, The Our Lady of The Lake Regional Medical Center, Baton Rouge, Louisiana, U.S.A.
[3]Dept. of Communication Sciences & Disorders, Louisiana State University, Baton Rouge, Louisiana, U.S.A.
[4]Dept. of Otorhinolaryngology Head & Neck Surgery, Medical School Division of Phoniatrics and Pediatric Audiology, University Hospital Erlangen, Erlangen, Germany
[5]Dept. of Speech, Language, & Hearing Sciences, University of Arizona, U.S.A.

Keywords: Voice onset/offset; Vowel-consonant-vowel; Highspeed videoendoscopy

Abstract

Objectives / Introduction:

The mechanics and kinematics of vocal fold vibration during the transients—either onset or offset of sustained phonation or in speech context—potentially contain rich information that can be related to vocal health, dysfunction, and patients specific voice symptoms. Measuring objective features of the transients from real speakers can be challenging, especially in speech context, because the non-transient segments are also dynamic with constantly fluctuating amplitude and frequency. Thus, there is no clear reference point to determine when the signal is out of the transient reaching so-called “steady state”.

In literature, the transient segment of sustained phonations were determined via model fitting[1], by a percentage of the steady-state amplitude[2], by the maximum glottal length[3], or by utilizing a feature such as vocal attack time[4] which does not depend on transient/steady-state boundary. Yet another approach, which has been explored in speech context, is the voiceless-consonant-induced fundamental frequency[5] which is measured from a fixed number of vocal cycles from a consonant-induced cessation. This presentation proposes a method to establish objectively the transient/steady-state boundary by fusing the vibration amplitude envelope and instantaneous fundamental frequency estimates with piecewise linear approximation. Use of a piecewise linear function is a simple and effective way to unify the handling of vocal transitions influenced by both voiced and voiceless consonants.

Methods:

Given a one-dimensional voice signal capturing a vowel-consonant-vowel (VCV) transitions, the goal is to identify when the consonant-influenced segment begins and ends (let t_{ss} to mark the boundary). We assume that an approximate location of the consonant is given. Also, the influence of a consonant is assumed to lower the vibration amplitude envelope and fundamental frequency f_o always.

The proposed process is performed in two steps. First, the presence of vocal cessation is detected to establish the least active reference points (t_{min}). If cessation is present, its starting and ending times are determined. If not, the time markers of the minimum amplitude and f_o are recorded instead. The cessation analysis fuses the information from both envelope and f_o contours. Also, the AC-component of the voice signal is used refine the $t_{\min}$ estimate by detecting the vanishing point of its cyclic behavior.

With the least-active reference points, the second step models the behaviors of the envelope and f_o contours from least active to most with a 2-segment piecewise linear function. The junction between the two segments are expected near the edge of consonant-influenced segment. The vowel-side segment is then extended till it intersects the contour to form the first candidate of the boundary (t_0). Some consonant induces overshooting behavior on the envelope or f_o. This transient characteristic is captured by fitting another 2-segment function is fitted the vowel starting from the nearest peak to t_0 on the vowel side away from the consonant. The first intersection with the contour is marked t_2. If the %overshoot metric is over a preset threshold (e.g., 20%), t_2 is chosen as t_{ss} else t_1 is selected. Boundaries of feature contours are individually processed.

Results:

To demonstrate, Fig. 1 shows analysis outcomes of the glottal area waveforms (GAWs) of nasoendoscopic highspeed videoendoscopic recordings at 4000 frames/second. These are VCV patterns with /i/ and alveolar consonants (/iti/, /idi/, /isi/, /izi/). The algorithm properly handled both with and without cessation.

Conclusions:

Piecewise linear function is an effective tool to describe changes in the behaviors of the amplitude or f_o behavior of vocal signals in consonant-influenced segment from neighboring vowel-only segment. This is achieved by having three simple degrees of freedom to model specific features (the overall slope of each segment and the transition time).

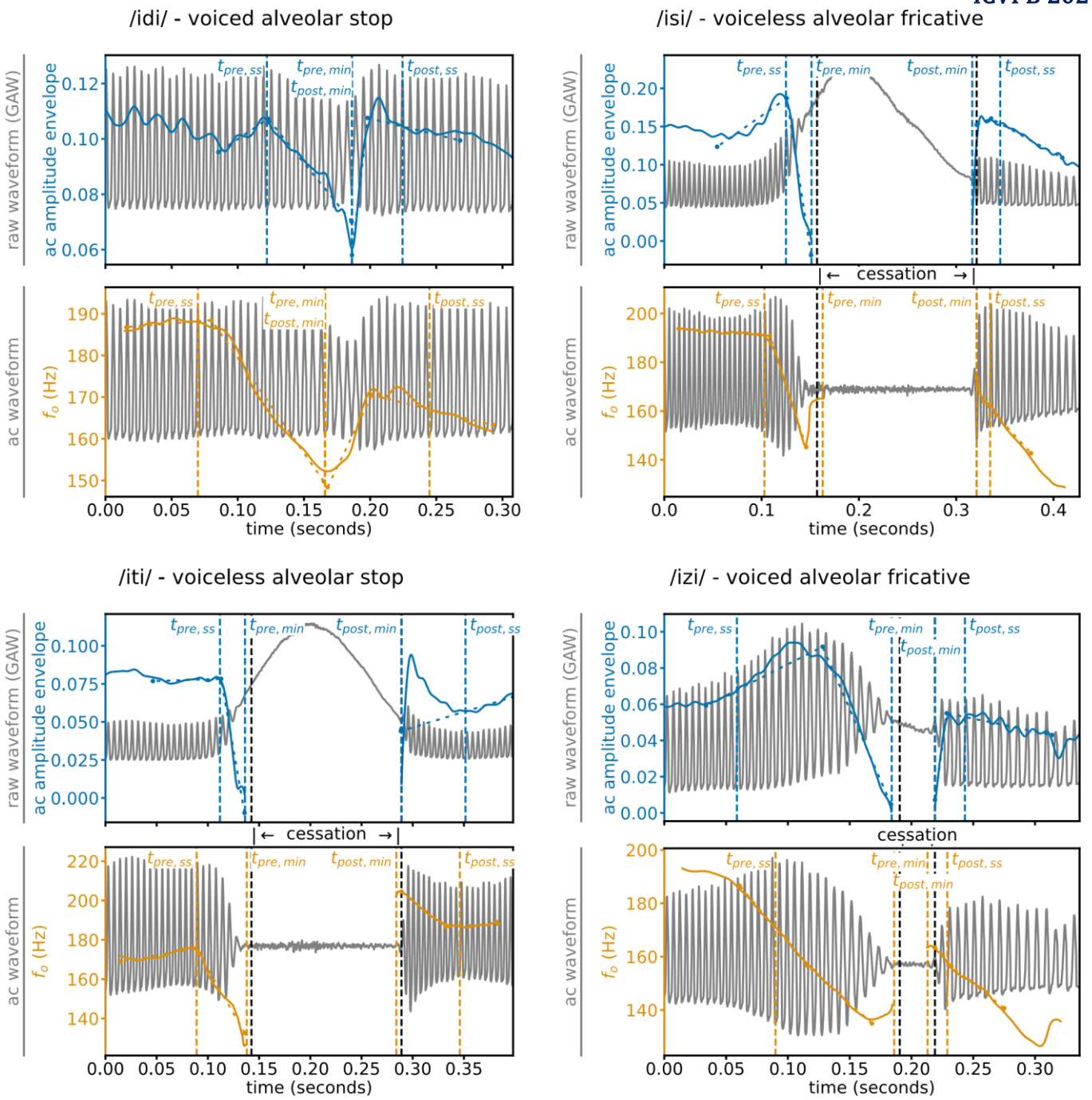


Fig. 1. Analysis outcome examples: Constant-influenced transient segments in alveolar VCV patterns of GAW signals.

References

1. Mergell P, Herzel H, Wittenberg T, Tigges M, Eysholdt U. Phonation onset: Vocal fold modeling and high-speed glottography. *J Acoust Soc Am*. 1998;104(1):464-470. doi:10.1121/1.423250

2. Kunduk M, Ikuma T, Blouin DC, McWhorter AJ. Effects of volume, pitch, and phonation type on oscillation initiation and termination phases investigated with high-speed videoendoscopy. *J Voice*. 2017;31(3):313-322. doi:10.1016/j.jvoice.2016.08.016

3. Ikuma T, Melda K, Fink D, McWhorter AJ. A spatiotemporal approach to the objective analysis of initiation and termination of vocal-fold oscillation with high-speed videoendoscopy. *J Voice*. 2016;30(6):756.e21-756.e30. doi:10.1016/j.jvoice.2015.09.007

4. Orlikoff RF, Deliyski DD, Baken RJ, Watson BC. Validation of a glottographic measure of vocal attack. *J Voice*. 2009;23(2):164-168. doi:10.1016/j.jvoice.2007.08.004

5. Löfqvist A, Baer T, McGarr NS, Story RS. The cricothyroid muscle in voicing control. *J Acoust Soc Am*. 1989;85(3):1314-1321. doi:10.1121/1.397462

THE SENSITIVITY OF PHONATION ONSET PRESSURE TO VOCAL FOLD SHAPE

Jonathan J. Deng[1], Sean D. Peterson[1]

[1] Department of Mechanical and Mechatronics Engineering, University of Waterloo, Waterloo, Ontario, Canada

***Keywords*: Finite element modelling, Phonation onset pressure, Vocal fold shape, Sensitivity analysis**

Abstract:

Objectives / Introduction: The vocal folds (VFs) possess widely varying shapes across individuals that can change for a given speaker with laryngeal posture. Onset pressure is an important characteristic of the ease of phonation and is affected by VF shape [Chan1997]. However, investigations into the effect of VF shape on onset pressure typically focus on a limited set of geometric parameters, such as the angle of the medial surface of an idealized fold [Lucero1998, Chan1997]. While these studies have successfully determined the influence of these specific parameters on the vibrational characteristics of an idealized fold, they are limited by the *a priori* selection of parameters. Furthermore, a broad systematic parametric exploration of the geometry is practically difficult given the relatively complex geometry of the VFs leading to a high-dimensional parameter space. Herein we use a finite element-based VF model where shape varies smoothly through deformations of the mesh to explore the effect of shape on onset pressure, an important characteristic of VF vibration. We use a sensitivity analysis to model how onset pressure changes with respect to any smooth stiffness change through a second order Taylor series approximation. To better interpret this model, we further decompose this Taylor series approximation, using a strategy similar to principal component analysis, into a small number of smooth shape modes that are primarily responsible for the variations in onset pressure. Unlike shape variations in parametric studies, these smooth shape modes are not defined *a priori* and so are not limited by a specific shape parameterization.

Methods: A two-dimensional (2D) finite-element (FE)-based VF model coupled with a 1D Bernoulli-based glottal flow model was used to determine the variation in phonation onset pressure with smooth shape variations. Phonation onset pressure was computed by identifying Hopf bifurcations of the coupled model [Titze1988, Zhang2007]. Specifically, a system of equations governing the conditions for a Hopf bifurcation [Griewank1983] of the coupled VF model were solved to identify the phonation onset pressure. Smooth variations in the shape of the VFs were treated by allowing the VF pre-phonatory shape to be deformed by an imaginary surface force. This strategy was employed to handle numerical difficulties with mesh motion since the interior mesh points should move to preserve mesh quality when the mesh surface shape is altered.

A second-order Taylor series approximation of the effect of shape changes from a base shape on onset pressure was then computed, such that changes in onset pressure are given by

$$p_{\mathrm{on}} \approx p_{\mathrm{on}}(\boldsymbol{m}_0) + \frac{\partial p_{\mathrm{on}}}{\partial \boldsymbol{t}}(\mathbf{m}_0)\boldsymbol{t} + \frac{1}{2}\boldsymbol{t}^{\mathrm{T}}\frac{\partial^2 p_{\mathrm{on}}}{\partial^2 \boldsymbol{t}}(\boldsymbol{m}_0)\boldsymbol{t},$$

where p_{on} is the onset pressure, $\boldsymbol{m}_0$ represents the base VF shape (mesh vertex locations), and $\boldsymbol{t}$ represents a shape change (an imaginary shape changing force). We computed the first and second-order sensitivities of onset pressure with respect to shape changes using an adjoint-based method. A ranking of shape changes that contribute the most to changes in onset pressure was then computed by an eigen-decomposition of the second-order sensitivity ($\partial^2 p_{\mathrm{on}}/\partial^2 \boldsymbol{t}$), similar to a principal component analysis.

To assess whether the results of this sensitivity analysis are general or simply a product of a specifically chosen VF base shape and base VF properties, we conducted the sensitivity analysis over a range of base VF parameters. We considered a range of base shapes centered around the M5 geometry [Scherer2001a] and a simple trapezoidal geometry over medial surface angles ranging from 0° to 30°. We also considered varying body to cover stiffness ratios of $1{:}\,1, 2{:}\,1, 3{:}\,1$, and $4{:}\,1$.

Results: To illustrate the results, we show in Figure 1 an example case for a base shape derived from the M5 model [Scherer2001] with a 3° medial angle for a $1{:}\,1$ body-to-cover stiffness ratio with nominal value of $6\ \mathrm{kPa}$. Figures 1A to 1C illustrate the first three shape modes ordered by decreasing eigenvalue magnitude. The shape mode with the biggest impact has an associated positive eigenvalue of $\lambda = 2072.8$ and consists of an inferior-superior lengthening of the VF, as well as rocking the vocal fold in the superior direction such that the glottal configuration is more divergent. The second shape mode consists of an inferior-superior lengthening with a depression near the midpoint of the medial surface and a lateral shrinking of the VF. The third shape mode consists of an inferior-superior lengthening of the VF with an increasingly convergent profile of the medial surface. These latter two shape modes represent more complex shape changes but have significantly smaller associated eigenvalues. This suggests that these shape modes contribute minimally to onset pressure variations in comparison to the first identified mode.

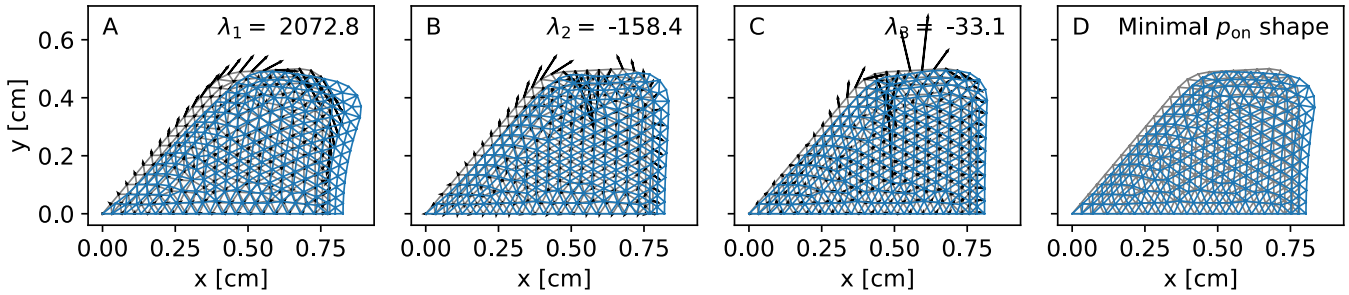


Figure 10: From left to right, the three shapes changes with the largest impacts on onset pressure, as ranked by eigenvalue magnitude, and a pre-phonatory shape that approximately minimizes onset pressure, as predicted by the sensitivity analysis. Arrows indicate changes in mesh vertex positions. The final image shows the original (grey) and modified vocal fold shape (blue) that minimized phonation onset pressure.

Because the first shape mode has an associated positive eigenvalue, this suggests that perturbing the VF shape by this mode results in quadratic increases in onset pressure, and therefore, a minimum in onset pressure can be achieved for a given perturbation by this mode. Based on the Taylor model, the perturbation along the first shape mode that minimizes onset pressure is shown in Figure 1D and involves a diverging glottal area and inferior-superior lengthening of the VF. The latter two shape modes with negative associated eigenvalues suggest the reverse of a positive eigenvalue; for a given perturbation along these shape modes onset pressure is decreased. Over the range of base geometries and body-to-cover stiffness ratios considered we found similar shape modes present. This suggests that the sensitivity of onset pressure to shape is not a product of a specific linearization point.

Conclusions: While there are infinitely many VF shape variations, the method introduced here enables identification of a set of shape modes that most affect onset pressure, ordered by significance of impact. In the present study, centered about M5-type vocal fold geometries, this shape change consists of a general inferior-superior lengthening of the VFs with rocking of the fold in the superior direction to create a more divergent glottis (see Figure 1). This shape change results in quadratic increases to onset pressure and, as a result, onset pressure can be minimized for a given scaling of this shape change. Our Taylor model suggests that a divergent glottal configuration is one contributor to minimizing onset pressure, which is consistent with past findings [Lucero1998]; however, we also find that a superior rotation of the VF along with inferior-superior lengthening also contribute to minimizing onset pressure. The full presentation will consider multiple iterations of the shape optimization process to explore the large deformations that may arise to minimize phonation onset pressure.

The main limitation of our work is the 2D VF geometry employed. The three-dimensional shape of real VFs would likely introduce anterior-posterior shape variations into the observed shape modes. What these anterior-posterior shape variations resemble will be the subject of future work.

Acknowledgements:

Research reported in this work was supported in part by the National Institute on Deafness and Other Communication Disorders of the National Institutes of Health under award P50DC015446.

References:

Griewank, A., & Reddien, G. (1983). The calculation of Hopf points by a direct method. IMA Journal of Numerical Analysis, 3(3), 295-303.

Lucero, J. C. (1998). Optimal glottal configuration for ease of phonation. Journal of Voice, 12(2), 151-158.

Titze, I. R. (1988). The physics of small-amplitude oscillation of the vocal folds. The Journal of the Acoustical Society of America, 83(4), 1536-1552.

Chan, R. W., Titze, I. R., & Titze, M. R. (1997). Further studies of phonation threshold pressure in a physical model of the vocal fold mucosa. The Journal of the Acoustical Society of America, 101(6), 3722-3727.

Scherer, R. C., Shinwari, D., De Witt, K. J., Zhang, C., Kucinschi, B. R., & Afjeh, A. A. (2001). Intraglottal pressure profiles for a symmetric and oblique glottis with a divergence angle of 10 degrees. The Journal of the Acoustical Society of America, 109(4), 1616-1630.

Zhang, Z., Neubauer, J., & Berry, D. A. (2007). Physical mechanisms of phonation onset: A linear stability analysis of an aeroelastic continuum model of phonation. The Journal of the Acoustical Society of America, 122(4), 2279-2295.

EXPLORING EFFICIENCY OF VOICE ACOUSTIC PARAMETERS FOR DETECTING VOICE DISORDERS: A SYSTEMATIC REVIEW OF LITERATURE

Ahmed M Yousef[1], Lady Catherine Cantor-Cutiva[1], Eric J Hunter[1]

[1]Department of Communication Sciences and Disorders, University of Iowa, Iowa City, Iowa, USA

***Keywords*: Voice Disorders; Voice Assessment; Machine Learning; Speech Acoustics**

Abstract:

Objectives / Introduction: Throughout the years, various voice acoustic measurements have been proposed and examined for their effectiveness in the assessment of speakers with voice disorders. The main aim of this paper is to determine the efficacy of different acoustic markers in evaluating voice disorders using a systematic review of literature.

Methods: A systematic review of literature was performed based on the PRISMA Statement guidelines. Indexed Journal and conference papers were collected from seven popular databases: Pubmed, Web of Science, Scopus, EBSCO (Academic Search Elite), Science Direct, BVS, and Scielo. In this search, papers centered on the assessment of voice disorders using voice acoustic parameters were considered, as well as two additional inclusion criteria, namely, papers conducted intra- or inter-comparison between dysphonic and non-dysphonic speakers and papers published in three different languages (English, Spanish, and Portuguese). In this ongoing study, the data from the included papers are analyzed through several types of analysis – bibliometric, co-occurrence, and content analysis. In addition, meta-analysis is performed to investigate the efficacy of specific acoustic measurements in assessment of different voice disorders.

Results and Conclusions: Our preliminary results suggest that the analysis of voice acoustic metrics and different types of voice problems and disorders has been widely performed. These acoustic markers are ranked in terms of their significant levels and correlation coefficients in assessing voice disorders. The outcome of the analysis is projected to also reveal the certain acoustic measurements that are most effective in assessing specific voice disorders. This review also highlights the potential measurements that need to be further investigated and the future directions for better acoustic assessment in clinical settings.

INFLUENCE OF SUPRAGLOTTAL TRACT ON FLOW SEPARATION VORTICES DURING VOCAL FOLD CLOSING IN HUMAN PHONATION

Weili Jiang[1], Liran Oren[2], Charles Farbos de Luzan[2], Ephraim Gutmark[3], Mahdi Sangbori[1], Xudong Zheng[1], Qian Xue[1]

1 Department of Mechanical Engineering, Rochester Institute of Technology, Rochester, NY, USA
2 Department of Otolaryngology-Head and Neck Surgery, University of Cincinnati, Cincinnati, OH, USA
3 Department of Aerospace Engineering, University of Cincinnati, Cincinnati, OH, USA

***Keywords*: Flow Separation Vortices; Supraglottal Tract; Intraglottal Negative Pressure**

Abstract:

Objectives / Introduction:

The oscillation of the vocal folds results from the interplay between glottal aerodynamics and tissue elasticity. In this process, intraglottal pressure serves as the driving force. This pressure during the closing phase of the fold vibrations is particularly interesting because of the vortical structures that can form by flow separation and impact the glottal dynamics [1,2]. Our previous study [3] showed through computational models that the mean negative pressure during glottal closing can be about 28% of the subglottal pressure. In the comparative cases where the negative intraglottal pressures were artificially removed, the vibration amplitude and flow rate were reduced by up to 20%, and the closing speed, flow skewness quotient, and maximum flow declination rate were reduced by up to 40%.

The main limitation of our previous study [3] was excluding a supraglottal tract from the model. Natural human phonation involves a supraglottal tract, and its presence affects the intraglottal pressure through flow inertia. In the current study, we aim to analyze the impact of the supraglottal tract on negative intraglottal pressures during glottal closing. We propose to examine two interactive mechanisms contributing to the negative intraglottal pressure formed during closing: (1) from the flow separation that forms in a divergent glottal shape and (2) from flow inertia in the supraglottal tract.

Methods:

Numerical simulation of glottal flow aerodynamics is conducted using a hydrodynamic/acoustic splitting method [4], which has been validated [5] and used in our previous work [3]. The simulation employs the three-mass model wherein the body-cover structure of the vocal fold is represented by three lumped masses connected through springs and dampers [6]. In the fluid-structure interaction (FSI) simulation, the displacement and velocity of the upper and lower masses from the three-mass model are provided to the flow model to update the flow field; additionally, the surface pressure from the flow model is utilized to calculate the external force on the three-mass model, therefore, driving its motion.

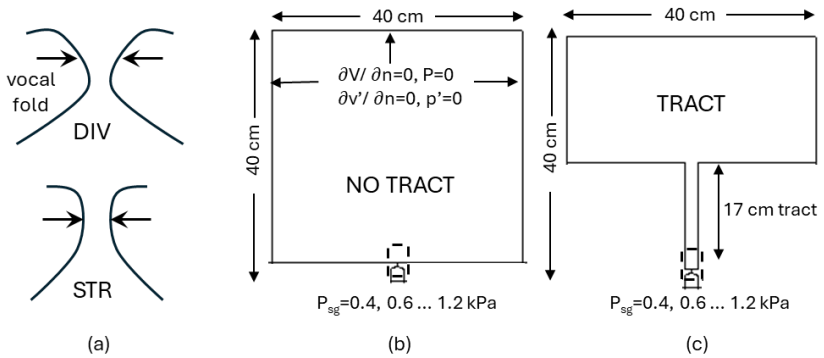


Figure 1. Simulation setup. (a) In the DIV (divergent) group, the glottis forms a divergent shape during closing. In the STR (straight) group, the glottis forms a straight channel during closing. These differences in glottal behavior are achieved by adjusting the parameters of the three-mass vocal fold model. (b) The open-space flow domain setup for NO TRACT condition. The glottis directly opens to a large open space (c) The flow domain setup for TRACT condition. The glottis is connected to a supraglottal tract, which opens to a large open space.

Four groups of simulations were conducted (Figure 1). In the DIV (divergent) groups, the three-mass parameters are consistent with those used in [3]. These parameters generate a divergent glottal shape during closing. The glottis is connected to an open space or supraglottal tract, representing the conditions without-tract (NO TRACT DIV) and withtract (TRACT DIV). In the STR (straight) groups, we adjusted the parameters of the three-mass model to achieve a

nearly straight channel shape during vocal fold closing in both without-tract (NO TRACT STR) and with-tract (TRACT STR) conditions. The open space and supraglottal tract geometries are identical for the DIV and STR groups. In each group, the subglottal pressure ranges from 0.4kPa to 1.2kPa with increments of 0.2kPa.

Results:

Sustained vibration was achieved in all cases. Figure 2 shows the minimum glottal angle, the peak glottal flow rate, and the mean pressure on the vocal fold medial surface during closing. The minimum glottal angle during vocal fold closing in the STR group is close to zero but can be slightly divergent. The peak flow rate was not affected by the tract but was higher when the vocal folds closed in a straight rather than divergent glottis shape.

Regarding the mean pressure during glottal closing, due to the minimal flow separation in STR cases, we propose that the effect of the supraglottal tract in the STR conditions is primarily due to flow inertia in the tract, while the impact of the supraglottal tract in the DIV conditions is due to the combined effects of flow inertia in tract and flow separation in the glottis. It is observed from Figure 2 that the presence of the supraglottal tract largely reduced intraglottal pressures in DIV cases. In contrast, this effect is much smaller or even the opposite in STR cases. Furthermore, it is noticed that the mean pressure difference is smaller between DIV and STR groups when no tract is present (No TRACT STR vs NO TRACT DIV in Figure 2), suggesting a smaller effect of flow separation on the pressures without a tract. These results suggest that the presence of the supraglottal tract can further reduce the negative intraglottal pressure during glottis closing by affecting flow separation-induced negative pressures, potentially by strengthening the intraglottal vortices. However, this hypothesis needs to be further examined.

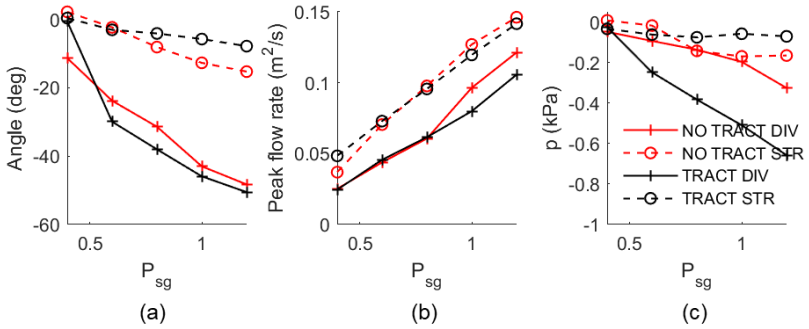


Figure 2. (a) Minimum glottal angle. (b) Peak glottal flow rate. (c) Mean pressure of the medial surface during vocal fold closing.

Conclusions:

During normal human phonation, the intraglottal negative pressure is affected by both the inertance of the supraglottal tract during flow declination and the flow separation in a divergent glottis. To quantify the impact of the supraglottal tract, we conducted two groups of simulations where the vocal folds closed either straightly or divergently, with and without a supraglottal tract. The results indicate that, in addition to directly reducing intraglottal pressure through the flow inertia in the tract, the supraglottal tract may also indirectly lower intraglottal pressure by reinforcing the negative pressure generated by flow separation.

Acknowledgements:

The research was funded by NIH Grant No. R01DC009435 from the National Institute on Deafness and Other Communication Disorders (NIDCD).

References:

1. Mihaescu, M.; Khosla, S.M.; Murugappan, S.; Gutmark, E.J. Unsteady Laryngeal Airflow Simulations of the Intra-Glottal Vortical Structures. J Acoust Soc Am 2010, 127, 435–444.
2. Khosla, S.; Murugappan, S.; Paniello, R.; Ying, J.; Gutmark, E. Role of Vortices in Voice Production: Normal versus Asymmetric Tension. Laryngoscope 2009, 119, 216–221.
3. Jiang, W.; Zheng, X.; Farbos de Luzan, C.; Oren, L.; Gutmark, E.; Xue, Q. The Effects of Negative Pressure Induced by Flow Separation Vortices on Vocal Fold Dynamics during Voice Production. Bioengineering 2023, 10, 1215.
4. Seo, J.H.; Mittal, R. A High-Order Immersed Boundary Method for Acoustic Wave Scattering and Low-Mach Number Flow-Induced Sound in Complex Geometries. J Comput Phys 2011, 230, 1000–1019.
5. Bodaghi, D.; Jiang, W.; Xue, Q.; Zheng, X. Effect of Supraglottal Acoustics on Fluid-Structure Interaction during Human Voice Production. J Biomech Eng 2021, 143.
6. Story, B.H.; Titze, I.R. Voice Simulation with a Body-Cover Model of the Vocal Folds. Journal of the Acoustical Society of America 1995, 97, 1249–1260.

REAL-TIME VOICE QUALITY ALTERATION: TACKLING THE COMPLEXITY OF SIMULATING DYSPHONIA

Isabel S. Schiller[1], Karolin Krüger[2], Tjorben Lerg[2], Patricia Weede[2], Gerhard Schmidt[2]

[1] Work and Engineering Psychology, RWTH Aachen University, Aachen, Germany
[2] Digital Signal Processing and System Theory, Department of Electrical and Information Engineering, Kiel University, Kiel, Germany

***Keywords*: Auditory Feedback Modulation, Voice Quality, Dysphonia, Voice Resynthesis**

Abstract:

The possibility of manipulating a talker's voice quality in real-time plays an important role in the context of auditory feedback modulation (AFM) and vocal motor control. AFM is a method in which a person phonates into a microphone and receives a direct acoustically perturbed feedback of their own voice via headphones. As these feedback modulations are processed, they typically result in the initiation of direct vocal responses. Our goal was to develop a system for real-time voice quality alteration that allows for the adjustment of a person's voice to authentically portray dysphonic attributes. In an iterative process, we manipulated different vocal parameters and added noise components to the spectrum to achieve a hoarse voice quality. Various resynthesis settings were tested in AFM experiments in which we collected perceptual ratings and assessed potential vocal changes through acoustic analyses. Here, we present and discuss our findings, offer a glimpse into future research directions, and examine potential implications for voice therapy.

Acknowledgements:

The project is funded under the Excellence Strategy of the Federal Government and the Länder (ref.: StUpPD_445-23). Isabel Schiller's contribution to this study was also funded by a grant from the HEAD-Genuit-Foundation (P-16/10-W) awarded to Prof. Sabine Schlittmeier. We acknowledge the support of our student assistants M. Sopha, M. Kretz, C. Baur, and S. Düwel.

NOTES ON CASCADING SIMPLIFICATIONS OF THE VOCAL TRACT AND ITS RELEVANCE FOR THE TRANSFER OF ACOUSTIC ENERGY

Mario Fleischer[1], Reuben Scott Walker[1,2] Peter Birkholz[3], Dirk Mürbe[1,2]

[1] Department of Audiology and Phoniatrics, Charité – Universitätsmedizin Berlin, Corporate Member of FreieUniversität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany

[2] Hochschule für Musik Carl Maria von Weber Dresden, Dresden, Germany

[3] Institute of Acoustics and Speech Communication, Technische Universität Dresden, Dresden, Germany

***Keywords*: Vocal tract; Finite-Element-Modeling; Acoustic analysis**

Abstract:

Objectives / Introduction:

The current gold standard for modeling the acoustic energy transfer of the vocal tract is the calculation of detailed 3dimensional models derived from a stack of MRI data [1]. It seems reasonable to simplify these complex geometries for increased convenience, reduced calculation time, and to limit the time needed to segment the entire cavity. Therefore, we derived several simplified models based on the 3-dimensional ones to study the impact of the simplifications on the acoustic transfer function.

Methods:

The first simplification comprises the removal of main side cavities, namely the piriform sinuses and the vallecula. Within a second approach, we derived cylindrical models with identical bent shapes (as the original model) and identical exact cross-sectional areas. In a third approach, we derived area functions from the cylindrical models. The detailed 3-dimensional models, the models without side cavities, and the cylindrical models were analyzed using the Finite-Element-Methods, whereas the area functions were analyzed within PRAAT [2].

Results:

We found that the elimination of the side cavities strongly changed the high-frequency transfer characteristics at frequencies higher than the second resonance frequency. For the cylindrical and the PRAAT models, we observed additional discrepancies in the first two resonance frequencies in comparison to the detailed 3-dimensional models.

Conclusions:

Depending on the degree of simplifications, the acoustical transfer characteristics of the vocal tract significantly change in terms of resonance frequencies and bandwidths. Despite characteristics sizes of side cavities and local pockets of a few millimeters up to a few centimeters (typical for wavelengths >> 10 kHz), these morphometric features also determine the frequency range for speech[3].

Acknowledgements:

We thank Pouriya Amini and Patrick Häsner for their help with model segmentation.

References:

[1] Birkholz, P.; Kürbis, S.; Stone, S.; Häsner, P.; Blandin, R. & Fleischer, M. Printable 3D vocal tract shapes from MRI data and their acoustic and aerodynamic properties *Scientific Data,* **2020***, 7*, 255

[2] Boersma, P. Praat, a system for doing phonetics by computer *Glot International,* **2001***, 5*, 341-345

[3] Arnela, M. & Ureña, D. Tuned two-dimensional vocal tracts with piriform fossae for the finite element simulationof vowels *Journal of Sound and Vibration,* **2022***, 537*, 117168

ACOUSTIC AND NEUROPHYSIOLOGICAL ASPECTS OF LOMBARD EFFECT

Lucia Z-Rivera *[1], Christian Castro *[1,2,3,4], Jhosmary Cuadros [1,5], Juan Pablo Cortés [1], Pavel Prado [6], Matías Zañartu [1,5]

[1]

[1]Advanced Center for Electrical and Electronic Engineering, Universidad Técnica Federico Santa María, Valparaíso, Chile

[2] Facultad de Ciencias de la Rehabilitación/Escuela de Fonoaudiología, Universidad Andres Bello, Santiago, Chile [3] Department of Speech and Hearing Sciences/Unit of Voice Disorders, Universidad de Valparaíso, Valparaiso, Chile.

[4] Department of Speech and Hearing Sciences/Unit of Voice Disorders, Universidad de Chile, Santiago, Chile.

[5] Department of Electronic Engineering, Universidad Técnica Federico Santa María, Valparaíso 2390123, Chile

[6] Escuela de Fonoaudiología, Facultad de Odontología y Ciencias de la Rehabilitación, Universidad San Sebastián, Santiago 7510602, Chile

***Keywords*: Lombard Effect; Electroencephalography; Dynamic Causal Modeling; Effective connectivity**

***First author**

Abstract:

Objectives / Introduction:

The Lombard Effect (LE) is the involuntary tendency of speakers to raise their voices in response to speaking in a loud or noisy environment (Bottalico et al., 2022). Essentially, LE acts as an adaptive communication mechanism and depends on the linguistic context of the message (Patel & Schell, 2008) and the type and intention of the communicative interaction (Garnier et al., 2010). The LE has been previously described through fMRI studies, where certain brain areas underlying the effect have been characterized (Meekings et al., 2016). This study aimed to compare variations in voice production and electrophysiological neural activity among individuals with healthy voices in different acoustic backgrounds. As such, this is the first approach to studying the LE using electroencephalography (EEG).

Methods:

Twenty-one volunteers were recruited for this study (mean age = 27.9, SD= 3.6). All participants completed three sequential acoustic background conditions: Baseline (in quiet), Lombard (in noise), and Recovery (in quiet after five minutes of rest). The acoustic signal was obtained using a microphone (B&K, model 4961; Nærum, Denmark) located in front of the participant at 15 cm from the lips at a 45-degree offset in the axial direction and amplified by a B&K 1705 signal conditioner. The EEG signals were recorded using a BioSemi ActiveTwo system with 64 active electrodes The location of each active electrode was according to the standard 10–20 montage. Dynamic Causal Modeling (DCM) was employed to investigate the effective connectivity involved in speech in noise.

Results:

Three DCM models were inverted for each subject's evoked responses of the Lombard Effect. Considering the source locations of the N1-P2 complex, the regions included in the models were the left primary auditory cortex (A1; [50 −10 8]), the left temporal pole (TPO; [55 10 -15]), the left parahippocampal gyrus (PHG; [30 −20 -15]), and the left inferior frontal gyrus (IFG; [50 30 4]). Specifically, in Model 1, the input was generated in A1; forward connections were specified from A1 to TPO, from TPO to IFG, and from TPO to PHG. Additionally, a backward connection from IFG to PHG was specified. Subsequently, a modulatory connection was outlined from PHG to TPO. In Model 2, the TPOPHG connection was replaced by the connection between A1 to PHG. In Model 3, there was specified a modulatory connection from PHG to A1, excluding TPO from modulation.

Bayesian model selection (BMS) analysis using random effects (RFX) inference method - at the group level - indicated exceedance probabilities of 97.16%, 1.52%, and 1.32% for Models 1, 2, and 3, respectively (Figure 1). Therefore, the probability of Model 1 being more likely than models 2 and 3 was higher. BMS was applied to determine the model that accurately fits the data. The comparison of models utilizes logarithmic evidence under the free energy criterion. This criterion consists of two elements: the precision (fit) term, obtained from the log-likelihood of the data, and the complexity term (Penny et al., 2004).

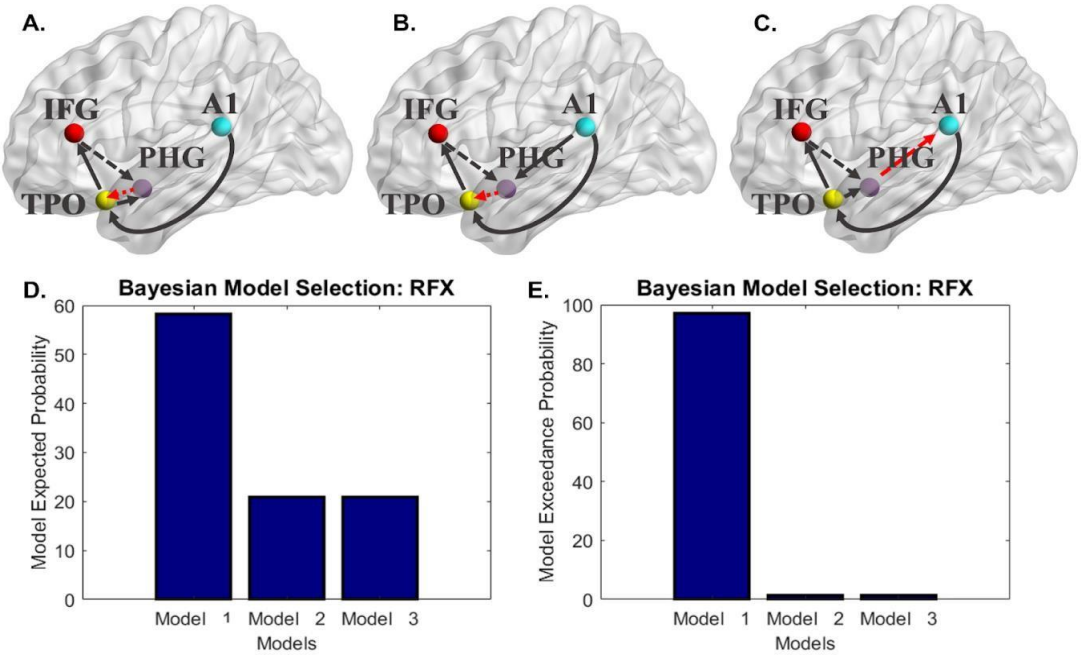


Figure 1. Three different models were constructed based on source localization analysis, incorporating the primary auditory cortex (A1), the temporal pole (TPO), the parahippocampal gyrus (PHG), and the inferior frontal gyrus (IFG). a) Model 1, b) Model 2, c) Model 3. The bar plots display d) the expected and e) exceedance probabilities for these three models.

Conclusions:

The findings indicate that forward, backward, and modulatory connections can explain the difference between the evoked response in the Lombard and baseline conditions. DCM inversion results provide insights into the cortical network underlying cognitive processes of prediction and feedback in the context of the LE. Furthermore, the modulatory effect of the connection from PHG to TPO offers guidance on the adaptive listening response in noisy environments within the LE.

Acknowledgements:

This research was supported in part by the National Institutes of Health (NIH) National Institute on Deafness and Other Communication Disorders grant P50 DC015446, and ANID FONDECYT 1230828 and BASAL FB0008.The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References:

Bottalico, P., Piper, R. N., & Legner, B. (2022). Lombard effect, intelligibility, ambient noise, and willingness to spend time and money in a restaurant amongst older adults. *Scientific Reports*, *12*(1), 6549. https://doi.org/10.1038/s41598-022-10414-6

Garnier, M., Henrich, N., & Dubois, D. (2010). Influence of Sound Immersion and Communicative Interaction on the Lombard Effect. *Journal of Speech, Language, and Hearing Research*, *53*(3), 588–608. https://doi.org/10.1044/1092-4388(2009/08-0138)

Meekings, S., Evans, S., Lavan, N., Boebinger, D., Krieger-Redwood, K., Cooke, M., & Scott, S. K. (2016). Distinct neural systems recruited when speech production is modulated by different masking sounds. *The Journal of the Acoustical Society of America*, *140*(1), 8–19. https://doi.org/10.1121/1.4948587

Patel, R., & Schell, K. W. (2008). The Influence of Linguistic Content on the Lombard Effect. *Journal of Speech, Language, and Hearing Research*, *51*(1), 209–220. https://doi.org/10.1044/1092-4388(2008/016)

Penny, W. D., Stephan, K. E., Mechelli, A., & Friston, K. J. (2004). Comparing dynamic causal models. *NeuroImage*, *22*(3), 1157–1172. https://doi.org/10.1016/j.neuroimage.2004.03.026

SPATIAL TRANSCRIPTOMICS OF THE VOCAL FOLD TISSUE OF INTACT AND OVARIECTOMIZED FEMALE RATS

Naila Cannes do Nascimento[1], Taylor W. Bailey[1], Abigail Cox[2], M. Preeti Sivasankar[1]

[1] Department of Speech, Language, and Hearing Sciences, Purdue University, West Lafayette, Indiana, USA
[2] Department of Comparative Pathobiology, Purdue University, West Lafayette, Indiana, USA

***Keywords*: Larynx, Vocal folds, Ovariectomy, Spatial Transcriptomics**

Abstract:

Objectives / Introduction: Sex hormones target many tissues, including the vocal folds (VF), making this organ a suitable candidate for studying hormonal impact in animal models that can be translated to humans. The rat vocal folds express estrogen receptor beta, emphasizing a functional but undetermined role for hormones in vocal fold biology[1,2]. As a multilayered organ composed of a mucosa (stratified squamous epithelium and extracellular matrix (ECM)-rich lamina propria) and thyroarytenoid muscle, a spatial investigation is crucial to interrogate the molecular signatures of each layer in response to estrogen manipulation. This preliminary study aimed to apply the GeoMx nanoString technology to compare the spatial transcriptional profile of each main layer of the rat VF - epithelium, lamina propria, and muscle - between intact and ovariectomized (OVX) rats.

Methods: Six-month-old intact (n=3) and ovariectomized (OVX, n=3) Sprague-Dawley female rats were used in this pilot study. The animals had no manipulation besides the ovariectomy surgery performed by the accredited vendor. Larynges were collected immediately post-euthanasia and placed in 10% formalin for FFPE blocks and subsequent spatial transcriptomics analysis. Two FFPE serial sections of the larynx were submitted for each rat (n=12 sections). After validation of morphological markers for each VF layer (epithelium (Ep), lamina propria (LP), and muscle (Ms)), a total of 72 regions of interest (ROIs) were selected across all FFPE slides, representing the Ep, LP, and Ms of the left and right VF, i.e., 4 ROI of each VF layer for each rat. Figure 1 shows an example of an OVX larynx section with ROIs selected from each tissue layer of the left and right VF. The spatial transcriptomics was performed using the GeoMx Mouse Whole Transcriptome Assay, which provides 13,668 probes with >90% identity to rat sequences, corresponding to >68% of the total mouse probes. Libraries were prepared for each ROI, and sequencing was performed using Illumina P2 100 cycles. Data were converted from FASTQ to DCC, and statistical analysis was performed by the nanoString statistical core service.

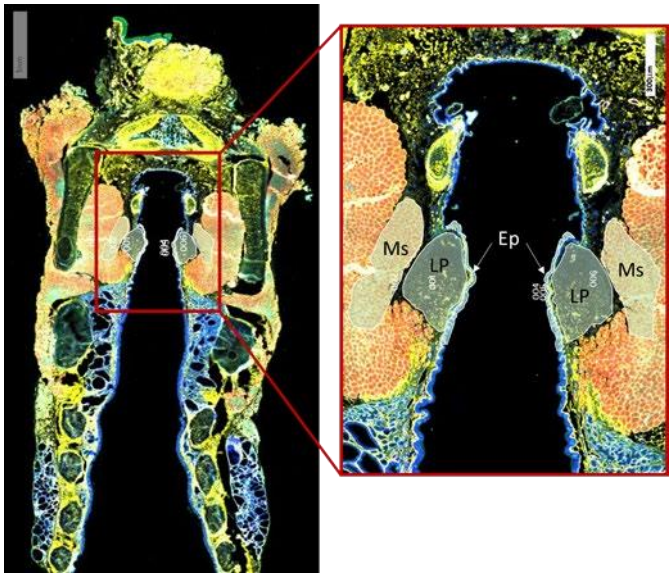


Figure 1: Representative figure of scanned rat larynx showing vocal fold tissue layers labeled with specific antibodies and selected regions of interest (ROIs).

Results: A total of 5,111 genes were identified in the rat VF mucosa and muscle. Principal component analysis (PCA) showed that samples are primarily clustered by tissue layer (figure 2). Despite the lack of statistically significant differentially expressed genes between intact and OVX rat larynges, the expression pattern was distinct between the tissue layers in each group, evidencing the different roles of each tissue to the vocal fold dynamics. This study was performed with a small sample size in order to determine the applicability of the GeoMx technology to rat VF tissue.

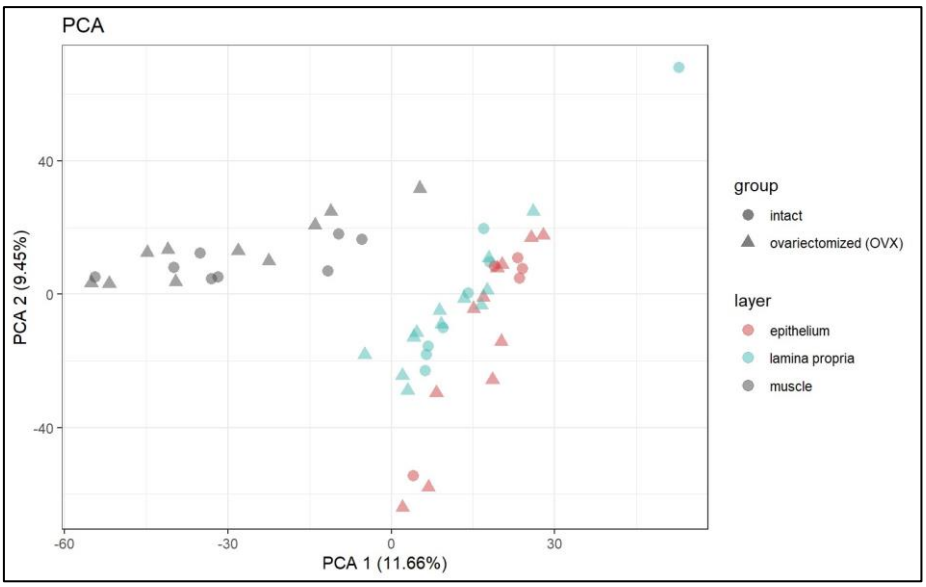


Figure 2: Principal component analysis showing clustering of vocal fold layers in intact and OVX rats, evidencing a primary separation by tissue layer regarding gene expression.

Conclusions: These preliminary results indicate that the GeoMx spatial transcriptomics can be applied to rat larynx tissue and that distinct gene expression patterns of vocal fold layers can be separated to uncover their differential role in vocal fold biology, especially when comparing epithelium and lamina propria to muscle. This study serves as a foundation for our current spatial transcriptomics of the rat larynx subjected to estradiol loss and replacement followed by systemic dehydration.

References:

1. Kim JM, Kim JH, Shin SC, et al. The protective effect of echinochrome a on extracellular matrix of vocal folds in ovariectomized rats. Mar Drugs. 2020;18(2). doi:10.3390/md18020077
2. Kim JM, Shin SC, Park GC, et al. Effect of sex hormones on extracellular matrix of lamina propria in rat vocal fold. Laryngoscope. 2020;130(3):732-740. doi:10.1002/lary.28086

THE INFLUENCE OF CIGARETTE SMOKE-STIMULATED VOCAL FOLD FIBROBLASTS ON INFLAMMATORY CYTOKINE RELEASE OF MACROPHAGES

Barbara Steffan[1], Magdalena Grill[1], Tanja Grossmann[1], Andrijana Kirsch[1,2], Markus Gugatschka[1]

[1] Department of Otorhinolaryngology, Division of Phoniatrics, Medical University of Graz, 8036 Graz, Austria
[2] current affiliation: Otto Loewi Research Center for Vascular Biology, Immunology and Inflammation, Division of Physiology and Pathophysiology, Medical University of Graz, 8010 Graz, Austria

***Keywords*: macrophages; fibroblasts; cigarette smoke; inflammation;**

Abstract:

Objectives / Introduction:

Macrophages are part of the immune regulation system and serve as an important cell type in dealing with the effects of various environmental harmful influences on the respiratory system such as toxic substances, and bacterial and viral infections (Hirayama, 2018). The IQOS™ system is a new smoking device that has been proposed to be a less toxic alternative to conventional cigarette smoking. This assumption was supported by chemical analyses from previously published research work (Uguna, 2022). However, because of the novelty of the development, the availability of data is scarce. The influence of other cell types on the cytokine profile of macrophages is an important research topic. Fibroblasts are known to build tissue structures, mediate signaling to neighboring tissue layers, and thus influence the microenvironment (Dutsch-Wicherek, 2013). We have previously investigated gene and protein expression levels of immortalized and primary fibroblasts from healthy and Reinke's edema patients and the altering effects of cigarette smoking as a major risk factor for several diseases. Current studies in our group focus on the synergistic effects of different cell types altered by cigarette smoke.

Methods:

In this preliminary study, we treated activated macrophages (M0) from the immortalized human THP-1 cell line with conventional cigarette smoke extract (CSE) and IQOS™ smoke extract (IQOS™-SE) for 24h. We compared the results with the effect of conditioned medium (CM) derived from CSE- and IQOS™-SE-treated immortalized human vocal fold fibroblasts (hVFF). Gene expression analysis was performed by RT-qPCR.

Results:

The monocyte chemotaxis receptor *CCR2* was downregulated by CSE, whereas CSE-CM was responsible for a significant upregulation.

Potential markers for CSE-induced cell damage (*CYP1B1*) and inflammatory genes (*COX2*, *NQO1*, *PPARγ*) were more upregulated by exposure to CSE than to IQOS™-SE. The slight increasing effect of IQOS™-SE was abolished with IQOS™-SE-CM, in fact a downregulation of *NQO1* and *PPARγ* was observed. This effect was not observed with CSECM.

COX1 and *MMP9* show similar downregulation with direct exposure to both smoke extracts and even stronger with CM treatment.

Conclusions:

These preliminary results showed that receptors (*CCR2* and *PPARγ*) and enzymes involved in smoking-induced cell damage (*CYP1B1*, *MMP9*) and acute inflammation (*COX1*, *COX2*, *NQO1*) are differentially regulated by molecular signals from hVFF. Our results indicate an enhanced effect on gene expression levels due to cross-talk of cell lines. We also noticed that conventional cigarette smoke extract had a stronger effect on macrophages than IQOS™-SE.

As a consequence, we suggest further exploration with co-cultivation of cell types to gain more detailed information for specific cell responses and also more accurate insight into cellular signaling across different cell lines.

Acknowledgements:

The work is supported by the Austrian Science Fund (FWF) grant P36067, awarded to M. Grill.

References:

Hirayama D, Iida T, Nakase H. The Phagocytic Function of Macrophage-Enforcing Innate Immunity and Tissue Homeostasis. International Journal of Molecular Sciences. 2018; 19(1):92. https://doi.org/10.3390/ijms19010092

Uguna, C. N., & Snape, C. E. (2022). Should IQOS Emissions Be Considered as Smoke and Harmful to Health? A Review of the Chemical Evidence. ACS Omega, 7(26), 22111–22124. https://doi.org/10.1021/acsomega.2c01527

Dutsch-Wicherek M, Kazmierczak W. Creation of a suppressive microenvironment by macrophages and cancerassociated fibroblasts. Front Biosci (Landmark Ed). 2013;18(3):1003-1016. Published 2013 Jun 1. https://doi.org/10.2741/4159

AMNIOTIC FLUID AS A POTENTIAL TREATMENT FOLLOWING VOCAL FOLD INJURY

Jenny Pierce[1], Brendan Olson[2], Ray Merrill[3], Anika Isom[4], Vanessa Torrecillas[1], Hilary McCrary[1], Marshall Smith[1], Ben Christensen[1,2]

[1] Department of Otolaryngology – Head & Neck Surgery, University of Utah School of Medicine, SLC, UT, USA
[2] Department or Surgery, Division of Urology, University of Utah School of Medicine, SLC, UT, USA
[3]Department of Communication Sciences & Disorders, Brigham Young University, Provo, UT, USA
[4]Department of Biology, University of Utah, SLC, UT, USA

***Keywords*: Vocal Fold; Scar; Inflammation; Elastic Modulus**

Abstract:

Objectives / Introduction: 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) has demonstrated healing potential in a variety of organs, including the airway. Therefore, we investigated the use of AF in a VF scar model.

Methods: This study included six groups of 10 (N=60) male New Zealand white rabbits. All groups included 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). Rheological analysis was conducted to determine viscous and elastic moduli, while RT-qPCR analysis was conducted to determine expression levels for the pro-inflammatory cytokines TNF-α, IL-1β, and IL6.

Results: Rheology: No significant differences in elastic modulus were detected between control VFs at week four or 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 than saline and HA groups. Differences within treatment and time groups between treated and non-injured VFs existed only in the saline group at four weeks. Differences within groups across time points existed in the saline non-injured group, and the AF treated group. Results for viscous modulus mirrored those for the elastic modulus.
PCR: For TNF-α, at four weeks, the saline group had a higher ΔCt than HA and AF groups. For IL-1β, differences were found in the saline control and AF control groups between four and 10 weeks, with the 10-week time points having higher ΔCt values. Diferrences were also found between treated and non-treated VFs in both the saline and AF groups at 10 weeks, with the control groups having higher values. Finally, ΔCt were higher at 10 weeks than at four weeks in the AF treated group. For IL-6, at 10 weeks, non-injured VFs in the HA group had higher values than treated VFs, and AF treated VFs had higher values than saline treated VFs.

Conclusions: Based on rheology results, AF treated VFs were different/stiffer than 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, with longer time points having lower inflammatory levels than shorter time points, and non-injured VFs having lower levels of inflammation than injured VFs. In general AF treated VFs at 10 weeks appeared to have lower levels of inflammation than saline treated VFs. These results suggest a potential for the use of AF as a treatment for VF scar, but further testing is needed.

Acknowledgements:

This work was funded by the Skaggs Research Foundation.

EXTRALARYNGEAL SURFACE EMG CLASSIFICATION DURING SENTENCE PRODUCTION FOR VOCAL FATIGUE DETECTION USING GA-SVM FOR CONFOUNDER REMOVAL

Y. Gao[1], G.N. DeSouza[1], M.L. Berardi[3], M. Dietrich[2,3]

[1] ViGIR Lab, University of Missouri, Columbia, MO, USA
[2] Department of Speech, Language and Hearing Sciences, University of Missouri, Columbia, MO, USA
[3] Department of Psychiatry and Psychotherapy, University Hospital Bonn, Bonn, Germany

***Keywords*: Vocal Fatigue; Machine Learning; surface EMG; Confounding Mitigation**

Abstract:

Introduction:

The objective of this work was to demonstrate the effectiveness of our classification via sEMG data collection of the extralaryngeal muscles to detect vocal fatigue using continuous speech while mitigating for confounding factors. Previously in [1] we proposed a classification framework where we extracted features from sEMG data collected on vowels and conducted subject-wise cross-validation to evaluate the overall classification performance. In this work, we applied the same confounding mitigation technique that was proposed in [2] to improve the generalization of our SVM approach applied to sentences.

Methods:

Table 1 shows our subject statistics and Table 2 shows the continuous speech tasks that were used for our classification experiments in this study. Our proposed classification framework and confounding mitigation scheme (GA-SVM) are illustrated in Figure 1, more details can be found in [1] and [2].

Table 1. Descriptive statistics for age, neck skinfold thickness (supra- and infrahyoid), and Vocal Fatigue Index factor 1 (VFI-1) scores for 40 matched test subjects.

	Vocally Fatigued	Vocally Healthy
Number of subjects	20	20
Age (21-39 years)	25.6 ± 4.3	25.3 ± 4.7
Suprahyoid (3.2-16.7 mm)	7.0 ± 3.4	5.4 ± 1.3
Infrahyoid (2.7-15.0 mm)	6.4 ± 3.1	5.1 ± 1.3
VFI-1 (0-28)	18.2 ± 5.4	5.1 ± 1.3

Table 2. Continuous speech tasks chosen for sEMG data collection used for the classification experiments.

Task	Description	Reps	Time
Sentence 1	“The dew shimmered over my shiny blue shell again” [3]	55	2 secs
Sentence 2	“Only we feel you do fail in new fallen dew” [3]	55	2 secs

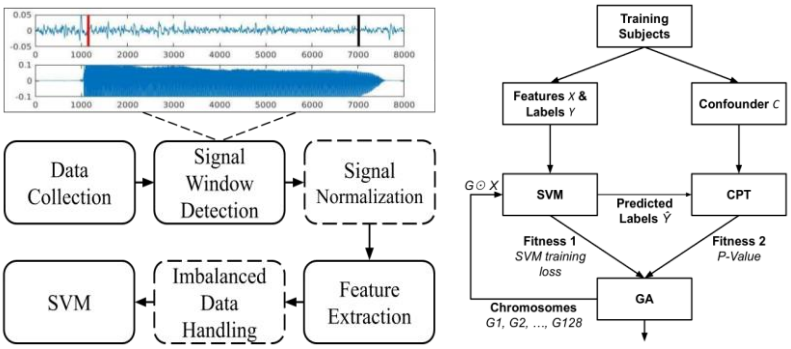


Figure 1. Proposed classification framework using Support Vector Machine (SVM) as the classifier and a Genetic Algorithm (GA)-based technique to mitigate confounding predictions biased by subjects’ skinfold thicknesses.

Results:

We used classification accuracies to evaluate our proposed framework. We also used *p*-values to measure whether our model predictions were confounded against subjects' skinfold thicknesses. The higher the *p*-value, the less confounded our model predictions were.

Table 3. Classification accuracy and *p*-value evaluations using subject-wise validation for sentences 1 (Table 3a) and 2 (Table 3b).

	p-value	Training Acc	Testing Acc
SVM	0.0058	98.61%	65.62%
GA-SVM	**0.2324**	**92.97%**	**90.23%**

(a) Sentence 1

	p-value	Training Acc	Testing Acc
SVM	0.0057	98.72%	66.76%
GA-SVM	**0.2179**	**94.42%**	**90.63%**

(b) Sentence 2

Conclusions:

In this work, we continued our investigation on the detection of vocal fatigue using a new classification approach (GA-SVM) for sEMG signals collected from subjects' anterior necks. We extended our experiments to continuous speech, instead of individual vowels, to improve ecological validity for a preclinical screening environment. We classified sEMG samples collected from a group of 40 matched subjects and achieved better results compared to our previous work in [2]. Moreover, our confounding mitigation technique using GA demonstrated improved effectiveness in generalizing the SVM classifier.

Acknowledgments:

Research reported in this publication was supported by the National Institute on Deafness and Other Communication Disorders of the National Institutes of Health under Award Number R15DC015335 and R01DC018026.

References:

1. 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*, *10*, 4335. https://doi.org/10.3390/app11104335

2. Gao, Y., DeSouza, G. N., Berardi, M., & Dietrich, M. (2023). Removal of confounding factors using GA-SVM feature adaptation: Application on detection of vocal fatigue thru sEMG classification. *2023 IEEE Congress on Evolutionary Computation (CEC)*. https://doi.org/10.1109/cec53210.2023.10253983

3. Lien, Y. A. S., Gattuccio, C. I., & Stepp, C. E. (2014). Effects of phonetic context on relative fundamental frequency. *Journal of Speech, Language, and Hearing Research*, *57*, 1259–67. https://doi.org/10.1044/2014_JSLHR-S-13-0158

DOES FORCED WHISPER HAVE AN IMPACT ON VOICE PARAMETERS?

Theresa Pilsl[1], Marie Köberlein[1], Michael Döllinger[2], Jonas Kirsch[1], Matthias Echternach[1]

[1] Division Phoniatrics and Pediatric Audiology, Department of Otolaryngology, Munich Uni-versity Hospital and Faculty of Medicine, Munich University (LMU)

[2] Division of Phoniatrics and Pediatric Audiology, Department of Otolaryngology Head & Neck Surgery, University Hospital Erlangen, Friedrich-Alexander-University Erlangen-Nuremberg

***Keywords*: whispering – loading – voice production**

Abstract:

Introduction: There has been an assumption that whispering may impact vocal function, leading to the widespread recommendation against its practice after phonosurgery[1,2]. However, the extent to which whispering affects vocal function and vocal fold oscillation patterns remains unclear.

Methods: 10 vocally healthy subjects (5 male, 5 female) were instructed to forcefully whisper a standardized text for 10 minutes at a sound level of 70 dB(A), measured at a microphone distance of 30 cm to the mouth. Prior to and following the whisper loading, the dysphonia severity index (DSI) was assessed. Simultaneously, recordings of high speed videolaryngoscopy (HSV), electroglottography, and audio signals during sustained phonation on the vowel /i/ (250 Hz for females and 125 Hz for males) were analyzed after segmentation of the HSV material.

Results: The pre-post analysis revealed only minor changes after the intervention. These changes included a rise in minimum intensity, an increase in the glottal area waveform-derived open quotient, and the glottal gap index, along with a decrease in relative average perturbation. However, no statistically significant changes were observed in the harmonic-to-noise-ratio, the glottal- to-noise-excitation-ratio, and the electroglottographic open quotient.

Conclusion: Overall, the study suggests that there are only small effects on vocal function following a forced whisper loading.

Acknowledgements:

Matthias Echternach's (DFG, grant 409/1-4) and Michael Döllinger's contributions were supported by Deutsche Forschungsgemeinschaft (DFG, grant DO1247/8-2).

References:

1. Rubin AD, Praneetvatakul V, Gherson S, Moyer CA, Sataloff RT. Laryngeal hyperfunction during whispering: reality or myth? J Voice 2006; 20:121-127.

2. Fleischer S, Hess MM. Postoperativer Stimmgebrauch: Ist Schweigen wirklich Gold? HNO Nachrichten 2018; 48:30-35.

THE PREVALENCE OF CREAK ACROSS BREATH GROUPS IN SPEAKERS WITH AND WITHOUT ADDUCTOR LARYNGEAL DYSTONIA

Katherine L. Marks[1], Saul A. Frankford [1], Sarah Cocroft [1], Samantha Lonergan[1], Cara E. Stepp[1,2,3]

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

***Keywords*: Voice; Acoustics; Laryngeal Dystonia, Creak**

Abstract:

Objectives / Introduction: Creak is an acoustic feature that has been used to discriminate speakers with adductor laryngeal dystonia (AdLD) from typical speakers during a standard reading passage with outstanding diagnostic accuracy [1]. Yet, creak is a phenomenon not unique to AdLD. In the linguistics literature, creak has been studied extensively in typical speakers as a phrase-boundary marker. This study aims to further investigate creak in speakers with AdLD and controls without voice complaints by comparing the prevalence of creak across estimated breath groups. Specifically, we hypothesized that creak would be consistent across breath groups in speakers with LD, whereas creak would be located at the end of phrases/breath groups in control speakers. Alternatively, we hypothesized that creak would be located toward the ends of estimated breath groups in both speaker groups; however, there would be more breath-like pauses in speakers with AdLD than controls, resulting in more creak overall. Secondarily, we hypothesized that the duration of breath-like pauses would be shorter in speakers with AdLD, leading them to run out of air at a faster rate than control speakers.

Methods: Thirty-four speakers were enrolled in the study, 17 of whom were diagnosed with AdLD and 17 of whom reported no current or history of voice complaints (controls). Each speaker was recorded reading aloud the first paragraph of the Rainbow Passage [2]. Two trained technicians then manually aligned a phoneme-specific transcription text grid to the acoustic signal in Praat [3]. Any pause > 500 ms was automatically marked as a breath-like pause. The technicians manually listened to pauses < 500 ms to determine whether a breath was audible, in which case it was marked accordingly. The technicians re-aligned and marked pauses for 20% of the recordings for the purposes of calculating intra- and inter-rater reliability. The presence of creak was estimated from the audio recordings in a time-varying way (10ms window) using a creak detection algorithm [4], implemented in a custom MATLAB [5] script. Creak occurrence on each phoneme was coded binarily depending on whether creak occurred. For each phoneme, the time between the phoneme midpoint and the onset of the next breath-like pause was calculated. A generalized linear mixed-effects model was performed to determine the relationship between the time preceding a breath-like pause and whether creak occurred while controlling for the effect of phoneme duration. Breath count and duration were also compared between groups to further examine whether these factors affect the prevalence of creak in AdLD.

Results: Inter-rater reliability between the two technicians was excellent (ICC(A,1) = .90), and intra-rater reliability was excellent (ICC (A,1) = .91 and .92). There was a greater overall probability of creak in the AdLD group compared to controls (22% vs. 5%, $p < 0.001$) and a greater probability of creak as speakers approached a breath-like pause in both groups ($p < 0.001$), consistent with the first hypothesis. Moreover, the interaction between time preceding breath-like pause and group was significant ($p < 0.001$), such that there was a stronger relationship between the time preceding a breath-like pause and creak for control speakers ($p < 0.001$). Overall, these results indicate that, within a breath group, creak is more consistent for speakers with AdLD than control speakers. Contrary to our secondary hypothesis, there were no significant differences in the total number of breaths between groups; however, speakers with AdLD took longer breath-like pauses ($p = 0.002$), consistent with an overall reduced rate of speaking ($p < 0.001$).

Conclusions: Creak is more prevalent in speakers with AdLD, despite there being no difference in the number of breaths taken between speakers with and without AdLD. The probability of creak occurring was greater in both groups the closer the phoneme was to the end of the breath group (i.e, the subsequent breath-like pause); however, differences in the slopes suggest that speakers without AdLD have a larger increase in the probability of creak occurring at the end of a breath group than speakers with AdLD. The shallower slope in speakers with AdLD indicates creak was more prevalent throughout the breath group than in controls. Results suggest that the underlying mechanisms that result in creak in speakers with AdLD may not only be related to breath groups, but possibly in response to laryngeal spasms. Further work is needed to investigate underlying causes of creak in speakers with AdLD.

Acknowledgements:

The authors acknowledge Daniel Buckley and Taylor Feaster with their assistance with data collection. The authors acknowledge Gregory Grillone, Pieter Noordzij, and Lauren Tracy for their assistance with patient recruitment and Jose Rojas Olavarria for his assistance with participant recruitment and coordination. This work was supported by grant DC015570 (C.E.S), DC020349 (K.L.M), and DC020598 (K.L.M) from the National Institute on Deafness and Other Communication Disorders, a Speech Science Grant (K.L.M) from the American Speech-Language-Hearing Foundation, and a Student Research Grant (K.L.M.) from Sargent College, Boston University.

References:

1. Marks, K.L., et al., *Automated Creak Differentiates Adductor Laryngeal Dystonia and Muscle Tension Dysphonia.* The Laryngoscope, 2023.
2. Fairbanks, G., *Voice and Articulation Drillbook*. Vol. 2. 1960, New York: Harper and Row.
3. *Praat: doing phonetics by computer[http://www.fon.hum.uva.nl/praat/]*.
4. Drugman, T., J. Kane, and C. Gobl, *Data-driven Detection and Analysis of the Patterns of Creaky Voice.* arXiv preprint arXiv:2006.00518, 2020.
5. MathWorks, *MATLAB*. 2021: Natick, MA.

List of Participants A to Z

Name	Institution	Email
Abur, Defne	University of Groningen	d.abur@rug.nl
Bui, Trung Kien	DR Inserm Paris Ilde france Centre Est	trungkien.bui@curie.fr
Bailly, Lucie	CNRS	lucie.bailly@3sr-grenoble.fr
Becker, Stefan	University Erlangen	stefan.becker@fau.de
Berardi, Mark	University Hospital Bonn	mark.berardi@ukbonn.de
Berry, David	UCLA	daberry@ucla.edu
Brockmann-Bauser, Meike	University Hospital Zurich	meike.brockmann-bauser@usz.ch
Calvache, Carlos	Corporación Universitaria Iberoamericana - Vocology Center	cabbeto@gmail.com
Cantor Cutiva, Lady Catherine	University of Iowa	ladycatherine-cantorcutiva@uiowa.edu
Cha, Junseo	University of Wisconsin-Madison	cha@surgery.wisc.edu
Christensen, Ben	University of Utah School of Medicine	ben.christensen@utah.edu
Ciucci, Michelle	University of Wisconsin	ciucci@surgery.wisc.edu
Contreras-Ruston, Francisco	Universitat de Barcelona, Maastricht University and Universidad de Valparaiso	francisco.contreras@uv.cl
Cortés, Juan	Universidad Técnica Federico Santa María	juan.cortess@usm.cl
Da Silva, Andrey	Federal University of Santa Catarina	andrey.rs@ufsc.br
De Paula Soares, Maria Francisca	Federal University of Bahia	mariafps@ufba.br
Deliyski, Dimitar	Michigan State University	ddd@msu.edu
Deng, Jonathan	University of Waterloo	j8deng@uwaterloo.ca
Dietrich, Maria	Universitätsklinikum Bonn	maria.dietrich@ukbonn.de
Dindart, Juliette	DR Inserm Paris Ilde france Centre Est	juliette.dindart@curie.fr
Dinesh, Chhetri	UCLA	DChhetri@mednet.ucla.edu
Dion, Gregory	University of Cincinnati	diongy@ucmail.uc.edu
Döllinger, Michael	University Hospital Erlangen	Michael.Doellinger@uk-erlangen.de
Donhauser, Jonas	Universitätsklinikum Erlangen	jonas.donhauser@uk-erlangen.de
Dragicevic, Daria	Boston University	ddragic@bu.edu
Echternach, Matthias	LMU Munich	matthias.echternach@med.uni-muenchen.de
Erath, Byron	Clarkson University	berath@clarkson.edu
Farbos de Luzan, Charles	University of Cincinnati	farboscs@ucmail.uc.edu
Fischer, Johannes	University Medical Center Freiburg	johannes.fischer@uniklinik-freiburg.de
Fleischer, Mario	Charité – Universitätsmedizin Berlin	mario.fleischer@charite.de
Foote, Alexander	UCSD	agfoote@health.ucsd.edu
Frouin, Frédérique	DR Inserm Paris Ilde france Centre Est	frederique.frouin@inserm.fr
Grill, Magdalena	Medical University of Graz	ma.grill@medunigraz.at
Groen, Swen	McGill	swen.groen@mail.mcgill.ca
Grossmann, Tanja	Medical University of Graz	tanja.grossmann@medunigraz.at
Gugatschka, Markus	Medical University Graz	markus.gugatschka@medunigraz.at

Heller Murray, Elizabeth	Temple University	liz.heller.murray@temple.edu
Heningson, Jan-Ole		
Henrich Bernardoni, Nathalie	CNRS	nathalie.henrich@gipsa-lab.fr
Herzel, Hanspeter	Charite	h.herzel@biologie.hu-berlin.de
Hinganu, Marius Valeriu	University of Medicine and Pharmacy "Grigore T. Popa"	marius.hinganu@umfiasi.ro
Hoffmeister, Jesse	University of Minnesota	hoff0692@umn.edu
Hoyer, Patrick	Fraunhofer	patrick.hoyer@zv.fraunhofer.de
Ikuma, Takeshi	Louisiana State University Health Sciences Center	tikuma@lsuhsc.edu
Jakubass, Bernhard	Michigan State University	bernhard-jakubass@gmx.de
Jordan, Paula	University Hospital Freiburg	paula.jordan@uniklinik-freiburg.de
Katz, Ute	University Hospital Erlangen	Ute.Katz@uk-erlangen.de
Khairuddin, Khairy Anuar Mohd		khairy@usm.my
Kist, Andreas	FAU Erlangen-Nürnberg	andreas.kist@fau.de
Krasko, Maryann	University of Wisconsin-Madison	krasko@surgery.wisc.edu
Kreimann, Jody	UCLA	jkreiman@ucla.edu
Kunduk, Melda	Louisiana State University	mkunduk@gmail.com
Kopp, Ann-Kathrin	University Hospital Erlangen	ann-Kathrin.Kopp@uk-erlangen.de
Köberlein, Marie	LMU Munich	marie.koeberlein@med.uni-muenchen.de
Kniesburges, Stefan	University Hospital Erlangen	Stefan.Kniesburges@uk-erlangen.de
Lagier, Aude	CHU de Liege	aude.lagier@chuliege.be
Lamprecht, Raphael	UMIT TIROL – Private University for Health Sciences and Health Technology	raphael.lamprecht@umit-tirol.at
Lehoux, Sarah	University of California, Los Angeles	slehoux@ucla.edu
Li-Jessen, Nicole	McGill University	nicole.li@mcgill.ca
Long, Jennifer	UCLA	JLong@mednet.ucla.edu
Lungova, Vlasta	UW Madison School of Medicine and Public Health	lungova@surgery.wisc.edu
Madill, Catherine	University of Sydney	cate.madill@sydney.edu.au
Marks, Katherine	Boston University	katherine.l.marks@gmail.com
Marscheider, Judith	Hospital rechts der Isar Munich	judith.marscheider@mri.tum.de
Martinez, Josue	Universidad Tecnica Federico Santa Maria	josue.martinez@usm.cl
Maxfield, Lynn	University of Utah	lynn.maxfield@utah.edu
Morton-Jones, Mariah	Auburn University	mariahem18@gmail.com
Nota, Yukiko	National Institute for Japanese Language and Linguistics	ynota@ninjal.ac.jp
Näger, Christoph	FAU Erlangen-Nürnberg	christoph.naeger@fau.de
Oren, Liran	University of Cincinnati	orenl@ucmail.uc.edu
Parra, Jesus	Universidad Técnica Federico Santa María	jesus.parrap@sansano.usm.cl
Patel, Rita	Indiana University	patelrir@iu.edu
Peschel, Benjamin	Universitätsklinikum Erlangen Hals-Nasen-Ohren-Klinik, Kopf- und Halschirurgie	benjamin.peschel@uk-erlangen.de
Peterson, Sean	University of Waterloo	peterson@uwaterloo.ca
Pilsl, Theresa	Munich University Hospital and Faculty of Medicine, Munich University (LMU)	theresamariapilsl@gmail.com

Pokorny, Florian	Medical University of Graz	florian.pokorny@medunigraz.at
Popeil, Lisa	Voiceworks	lisapopeil@mac.com
Radolf, Vojtech	Institute of Thermomechanics of the CAS	radolf@it.cas.cz
Riede, Tobias	Midwestern University	triede@midwestern.edu
Samlan, Robin	University of Arizona	rsamlan@email.arizona.edu
Scheer, Aurelius	Friedrich-Alexander-Universität Erlangen-Nürnberg	aurelius.scheer@fau.de
Scheible, Florian	UMIT TIROL – Private University for Health Sciences and Health Technology	florian.scheible@umit-tirol.at
Schelhorn, Tony	Universitätsklinikum Erlangen	tony.schelhorn@uk-erlangen.de
Schiller, Isabel	RWTH Aachen University	isabel.schiller@psych.rwth-aachen.de
Schlegel, Patrick	Formerly: UCLA, Los Angeles	patrickschlegel93@outlook.de
Schlicht, Samuel	Institute of Polymer Technology, Friedrich-Alexander-Universität Erlangen-Nürnberg	samuel.schlicht@fau.de
Schoder, Stefan	TU Graz	stefan.schoder@tugraz.at
Schraut, Tobias	Universitätsklinik Erlangen	tobias.schraut@uk-erlangen.de
Schwabe, Emma	FAU	mail@emmaschwabe.com
Semmler, Marion	University Hospital Erlangen	marion.semmler@uk-erlangen.de
Sivasankar, Preeti	Purdue University	preeti@purdue.edu
Steffan, Barbara	Medical University of Graz	barbara.steffan@medunigraz.at
Stritt, Fiona	Freiburger Institut für Musikermedizin (FIM)	fiona.stritt@uniklinik-freiburg.de
Sutor, Alexander	UMIT TIROL	alexander.sutor@web.de
Svec, Jan G.	Faculty of Science, Palacky University	jan.svec@upol.cz
Takahashi, Jun	Osaka University of Arts Junior College	johnandiamo@gmail.com
Takemoto, Hironori	Chiba Institute of Technology	hironori.takemoto@p.chibakoudai.jp
Ternström, Sten	KTH Royal Institute of Technology	stern@kth.se
Thibeault, Susan	University of Wisconsin Madison	thibeault@surgery.wisc.edu
Titze, Ingo	University of Utah	ingo.titze@utah.edu
Tlaidi, Mounib	Gipsa lab	mounib.tlaidi@gmail.com
Toda, Natsuki	Chiba Institute of Technology	s17c3087mt@s.chibakoudai.jp
Tracicaru, Rares	"Gr. T. Popa" University of Medicine and Pharmacy	trares2006@gmail.com
Traser, Louisa	Medical Center, University of Freiburg, Freiburg, Germany	louisa.traser@uniklinik-freiburg.de
Tur, Bogac	University Hospital Erlangen	bogac.tur@uk-erlangen.de
Venkatraman, Anumitha	University of Wisconsin Madison	venkatraman@surgery.wisc.edu
Vojtech, Jenny	Boston University	jmvo@bu.edu
Walker, Reuben	Charité - Universitätsmedizin Berlin	reuben-scott.walker@charite.de
Willis, Emma	MGH Institute of Health Professions	ewillis5@mgh.harvard.edu
Wu, Liang	Department of Biomedical Engineering, School of Life Science and Technology	liangwu@xjtu.edu.cn
Xue, Qian	Rochester Institute of Technology	qxxeme@rit.edu

Yang, Xinyi	Xi'an Jiaotong University	yangxinyi@stu.xjtu.edu.cn
Yoshinaga, Tsukasa	Graduate School of Engineering Science, Osaka University	yoshinaga.tsukasa.es@osaka-u.ac.jp
Yousef, Ahmed	University of Iowa	aysef@uiowa.edu
Zanartu, Matias	Universidad Tecnica Federico Santa Maria	matias.zanartu@usm.cl
Zhang, Zhaoyan	UCLA	zyzhang@ucla.edu
Zheng, Xudong	Rochester Institute of Technology	xxzeme@rit.edu