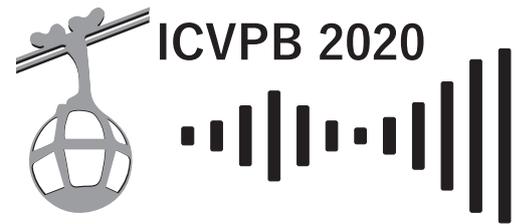Celebrating **40 years** of scientific advances
in **Vocal Fold Physiology** and **Biomechanics**

2-4th dec 2020
an Online event

ICVPB 2020

PROCEEDINGS   DECEMBER 2020

gipsa-lab

3SR
Laboratoire
Sols ▾ Solides ▾ Structures ▾ Risques

Sfa

©Jérome NARCY

Dear Colleagues,

On behalf of the local organizing committee, we are very pleased to welcome you to the 12th edition of the International Conference on Voice Physiology and Biomechanics ICVPB2020 (icvpb2020.sciencesconf.org). Together, we celebrate the 40th anniversary of scientific advances in Vocal Fold Physiology and Biomechanics !

ICVPB dates back to 1980. Initially called the Vocal Fold Physiology Conference, it began with five inspired pioneers who brought together voice scientists from Japan and the United States : Wilbur James Gould, Osamu Fujimura, Kenneth Stevens, Minoru Hirano, and Ingo Titze. The first meeting was held in Kurume Japan, in 1980. The focus was basic science, the physical and biological underpinnings of voice production. In total, nine Vocal Fold Physiology meetings were held. After Kurume, the meeting took place in Madison (1982), Iowa City (1984), New Haven (1985), Tokyo (1987), Stockholm (1990), Denver (1992), Kurume (1994), and Sydney (1995). The name of the conference was then changed to ICVPB to include the influx of biomechanics and biology into the study of voice production. The first ICVPB meeting was held in 1997 at Northwestern University in Evanston, Illinois, followed by Berlin (1999), Denver (2002), Marseille (2004), Tokyo (2006), Tampere (2008), Madison (2010), Erlangen (2012), Salt Lake City (2014), Viña del Mar (2016), East Lansing (2018), and now in Grenoble (2020), a place renowned in France for his long history of research in experimental phonetics. In particular, the Institut de Phonétique de Grenoble was founded in 1904 under the impetus of Théodore Rosset, a brilliant linguist and phonetician.

It is now a great honor for us to welcome this very special edition of ICVPB 2020. The event is the initiative of two Grenoble laboratories: the GIPSA-lab and the 3SR Lab. GIPSA-lab (Grenoble Images Speech Signal and Control) is a joint research unit of Univ. Grenoble Alpes, CNRS and Grenoble INP, which has recognized expertise in characterization and modeling of human voice. The 3SR Lab (Soils, Solids, Structures, Risks) is a joint research unit of Univ. Grenoble Alpes, CNRS and Grenoble INP, which has recognized expertise in the mechanics of (bio)materials and structures. ICVPB2020 has been organized with the team leading the Speech Acoustics Group (GAP) of the French Acoustical Society (SFA).

ICVPB2020 was scheduled to take place from 16th to 20th March 2020 (ICVPB2020 March Edition). But that was without counting on this terrible and unpredictable event that hit the world in the beginning of Year 2020, closing borders and restricting travel. A novel coronavirus (COVID-19) causing acute illness with severe respiratory symptoms has appeared and spread around the world. In France, the first major containment of the population was imposed by the French government from 17th March to 11th May 2020 as a health response to stop the spread of COVID-19 and limit its impact. ICVPB2020 was postponed to December 2020 and the two days of pre-courses scheduled prior to the main conference days were cancelled. We were hoping for a hybrid conference, both on site and online. However, a second mass containment of

French population started on October 30th to fight a surging second wave of the Covid-19 epidemic. ICVPB2020 is the first International Conference in Voice Physiology and Biomechanics to be held entirely online.

The ICVPB2020 scientific program includes oral presentations in mono-session and five keynotes, as well as eight parallel poster sessions. Oral presentations are organized into six thematic sessions, including:

- ◇ Session #1: Vocal fold tissue properties : biomechanics, biology and biomimicry

- ◇ Session #2: Flow-induced vibrations : modelling, simulation and validation

- ◇ Session #3: Singing

- ◇ Session #4: Voice analysis and synthesis

- ◇ Session #5: Voice disorders and treatments

- ◇ Session #6: Fluid-structure-sound interaction in the vocal tract

We would like to acknowledge all the people involved for helping us build a very rich conference program. More specifically, we are most grateful to the five keynote and plenary speakers for accepting our invitation, Drs. Ingo R. Titze, Brad Story, Sten Ternström, Caroline Boudoux and Jan G. Švec; all the authors for submitting the abstracts; the Scientific Committee for managing the reviewing process; tutors of the cancelled pre-courses days; all the chairmen/women for the moderation of the sessions; the ICVPB international advisory board for their precious advice; the SFA-GAP Bureau members and the local staff for all logistical aspects. We could not have done without their help !

We also wish to thank all our sponsors, whose support is essential for the success of this event: the laboratories and the SFA as co-organisers, the hosting institutions (CNRS, Univ. Grenoble Alpes, Grenoble INP), the Galileo Galilei Federation (Fed3G), the MACI, the Auvergne-Rhônes-Alpes Region, the French Society of Biomechanics (SB), the French-speaking Speech Communication Association (AFCP), the French Society of Phoniatrics and Laryngology (SFPL), and our industrial partners, NOVITOM, INSTRON and ZAION.

Finally, we would like to thank all the participants for their support and nice messages we received after rescheduling the March 2020 Edition, due to sanitary crisis. Thank you very much for your confidence, your patience during these last months and your presence to this December Edition !

Yours sincerely,

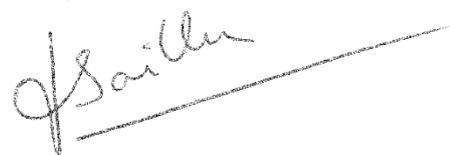Nathalie Henrich Bernardoni   and   Lucie Bailly

Table of contents

Session #3 - Singing 17

Session #4 - Voice analysis and synthesis 24

Session #1 - Vocal fold tissue properties : biomechanics, biology, biomimicry

Frequency analysis of ex-vivo porcine vocal fold elasticity using pipette aspiration

Florian Scheible[1]*, Raphael Lamprecht[1], Marion Semmler[2], Michael Döllinger[2] and Alexander Sutor[1]

[1]Institute of Measurement and Sensor Technology, UMIT - Private University, Hall i.T., Austria
[2]Div. of Phoniatrics & Pediatric Audiology, Dep. of Otorhinolaryngology, University Hospital Erlangen, Germany

Keywords: elastic modulus; vocal folds; pipette aspiration; frequency dependency

Introduction

To understand the complex process of phonation, it is necessary to know not only the static material properties of the tissue, but also the frequency depended.

In material properties studies of vocal folds pipette aspiration is a common tool, for dynamic measurements to observe its variation with frequency [1-3].

Methods

Figure 1 shows the measurement setup with the used pipette, which is pressed at the middle of the antero-postero line of one porcine vocal fold (VF), with a certain force, estimated by a load cell. A speaker produced sinusoidal pressure exciting the tissue and its velocity is determined by a vibrometer laser beam that permeates through the pipette.Conclusively, the absolute deformation is calculated by integration. A calibrated microphone is measuring the exact pressure.
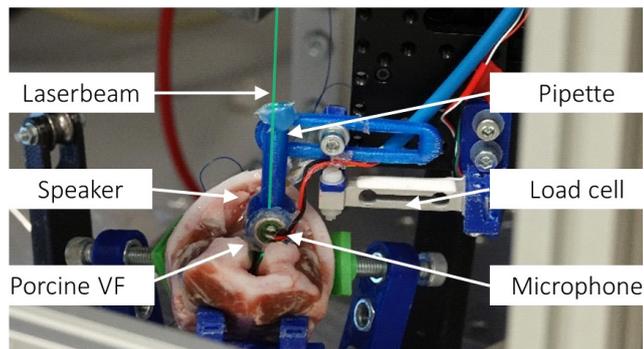
Figure 1: Measurement setup on porcine VF. Pressure produced by a speaker forces the sample to oscillate. The movement is detected by a laser vibrometer, as well as the pressure difference by a microphone.

Frequencies from 150 Hz to 1 kHz were produced with a step width of 50 Hz, and three different pressure amplitudes, each increasing up to 9, 14 and 19 Pa. Whereas each iteration is repeated six times. The pressure and velocity peak height of each frequency is determinate through a FFT over the signal length of one second. The measurement process is repeated after 13 min and 20 h to study the impact of temperature and humidity on the elasticity.The elasticity E is calculated with the following formula the pressure difference Δp is devided by the absolute displacement Δd

$$E = C \frac{\Delta p}{\frac{\Delta d}{di}}, \qquad (1)$$

with the pipette diameter di and a coefficient C dependent on the wall thickness [4].

Results

In Figure 2 the elasticity over the frequency for each pressure amplitude is shown as crosses. For each measurement (initial, after 13 min, after 20h) the mean value of the elasticity between the pressure amplitudes is plotted.
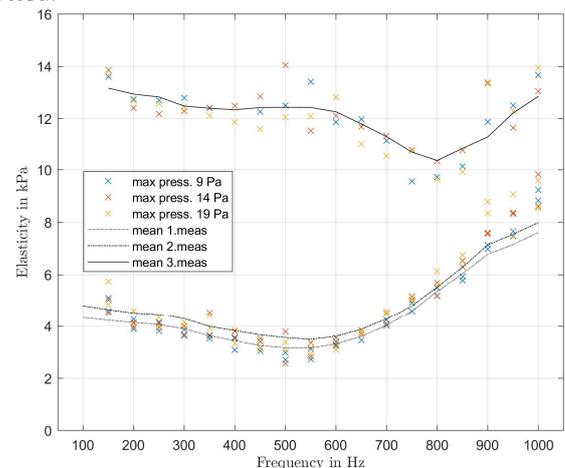
Figure 2: Elasticity with three different pressure amplitudes for each measurement (x), and the mean values (lines).

The first two measurements show the same characteristic, but the second measurement shows an increased elasticity E 8.26% (average over all Frequency) which can be explained by tissue dehydration. The third measurement after 20 h shows a different behavior with much higher values.

Discussion

To our knowledge, until now there are no ex-vivo measurements, to which our measurements could be validated.

Further investigations are planned to study variance in natural tissue and the effect of temperature and humidity of the tissue. Like previous investigations at our institute the frequency dependency of the elasticity will be determined with inverse numerical simulations [3].

Acknowledgements

References

[1] S. Weiss et al, *ACTA ACUST*, 102:214-229, 2016.
[2] S. Zörner et al, *Journal of Biomechanics*, 43:1540-1545, 2010.
[3] M. Maghzinajafabadi et al, *Applied Sciences*, 9:3875, 2019.
[4] T. Aoki et al, *ANN BIOMED ENG*, 25:581-587, 1997.

Strain measurements in ex-vivo porcine vocal folds using DeepFlow

Raphael Lamprecht[1]*, Florian Scheible[1], Marion Semmler[2], Michael Döllinger[2] and Alexander Sutor[1]

[1]Institute of Measurement and Sensor Technology, UMIT - Private University, Hall i.T., Austria
[2]Div. of Phoniatrics & Pediatric Audiology, Dep. of Otorhinolaryngology, University Hospital Erlangen, Germany
Keywords: vocal folds; ultrasound; strain; quasi-static

Introduction

The biomechanical properties of the true vocal folds play a crucial role in the phonation process. An ultrasound transducer directly placed on the vocal folds can measure the elastic properties by a quasi-static elastography approach. We already showed the feasibility of this technique in Lamprecht et al. [1]. Here we present an improved version of this approach.

Methods

An ultrasound transducer using a linear drive compressed the tissue of porcine vocal folds. To estimate the compressing force, a silicon pad placed between the transducer and the tissue (Fig. 1). The boundary between the tissue and the silicone was tracked by a Canny edge detection. Consequently, the acting force can be calculated with a lateral spatial resolution.
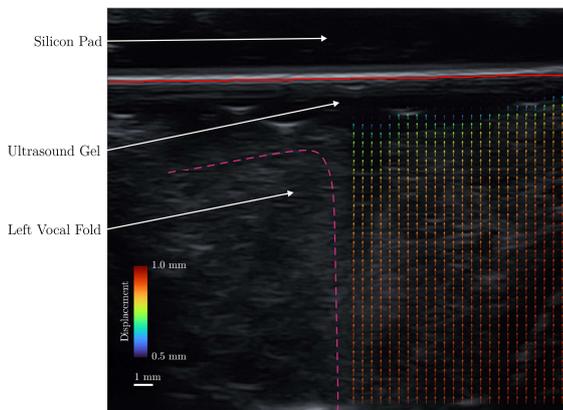
Figure 1: The vocal folds displayed in a coronary section. Left side: porcine vocal fold. Right side: Measured displacement in the tissue.

The displacement in the tissue was evaluated by DeepFlow, which is a trained neural network for optical flow estimation [2]. Using this, we were able to measure the displacement between the unconstrained and the loaded frame. The strain in the tissue was computed by a two-dimensional Savitzky-Golay differentiator [3, 4].

Results

We used two porcine larynges to test our algorithm. The displacement vector field in the tissue is shown in Fig. 1, whereas the color indicates the magnitude of the displacement. In the same figure, the border of the silicon pad is shown. Furthermore, Fig. 2 indicates the strain in the tissue in a user-defined region of interest. The color represents magnitude of the strain ε_{zz}, in which z is the direction of the compression.
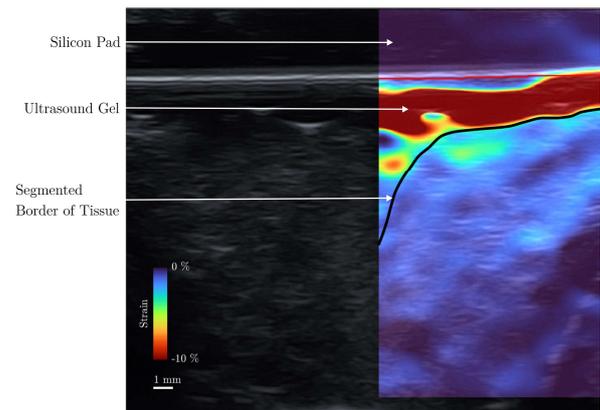
Figure 2: Strain ε_{zz} in the right vocal fold.

Discussion

We were able to show the feasibility of strain estimation in the vocal folds with DeepFlow, which is a major improvement of our previously presented approach [1]. In the future, a finite element model based on the body-cover-model [5] will be implemented to estimate the Young's Modulus of the tissue.

Acknowledgements

References

[1] Lamprecht, R. et al, *Applied Sciences*, 9(13):2729, 2019.
[2] Weinzaepfel, P. et al, *ICCV - IEEE International Conference on Computer Vision,* 1385-1392, 2013.
[3] Lai, W.M. et al, Introduction to Continuum Mechanics, 3rd ed., Oxford, NY, 1993.
[4] Pan, B. et al, *Optical Engineering*, 46(3):033601:1-033601:10, 2007.
[5] Hirano, M., *Folia Phoniatr,* 26(2):89-94, 1974.

*raphael.lamprecht@umit.at

Fabrication of synthetic, multi-material vocal fold models via embedded 3D printing

Taylor E. Greenwood, Scott L. Thomson*

Department of Mechanical Engineering, Brigham Young University, Provo, Utah, USA

Keywords: synthetic vocal fold models; 3D printing; embedded printing

Introduction

Synthetic, self-oscillating vocal fold (VF) models are physical models made from silicone rubber whose life-like vibration is induced and sustained by fluid flow. Among synthetic models, multi-layer models have been found to reasonably replicate the vibratory characteristics of human VFs [1]. Multi-layer models consist of multiple layers of silicone rubber, analogous to tissue layers within human VFs, where geometry and stiffness of each layer contribute to model performance. During fabrication, each layer is cast around previous layers in a multi-step process using a coordinated set of negative molds.

The development and use of multi-layer models, however, are limited because the iterative casting process is challenging and time intensive. A change in model geometry, for instance, requires a new set of positive and negative molds. Furthermore, fabrication yield is low because the interconnected nature of the material layers means an error in any one step often requires the casting process to be restarted.

The purpose of this study was to extend a recently-developed 3D printing process to the fabrication of multi-layer VF models. In addition to avoiding some of the challenges associated with the iterative casting process, we hypothesized that 3D printing would be faster, more material efficient, and potentially more versatile than casting, especially when creating models with new geometries.

Methods

An embedded 3D (EMB3D) [2] printing process was selected and adapted for fabricating the "EPI" VF model originally developed by Murray and Thomson [1,3]. This model consists of four material layers representing the body, ligament, superficial lamina propria (SLP), and epithelium. EMB3D printing is a hybrid of casting and 3D printing processes. In this method, a reservoir of the desired shape is filled with a curable gel-like material ("support matrix"). Shapes of different materials are then "embedded" within the support matrix by printing secondary materials throughout the matrix. In the present work we fabricated the EPI model by 3D printing the ligament and body layers within a single negative SLP mold as shown in Fig. 1.
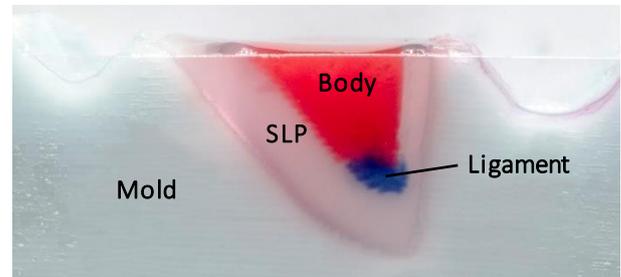
Figure 1: Side view image of an embedded 3D-printed VF model before demolding. The cavity in which the VF model is here seen was first filled with uncured SLP material. The ligament and body layers were then sequentially printed within the uncured material.

Results and Discussion

In this presentation, we will describe vibration testing performed on the printed models in hemilarynx and full larynx configurations to collect vibration frequency, onset pressure, amplitude, and flow rate data. Analyses of the geometry and material properties of the printed layers to determine the accuracy of 3D printing in replicating the desired model parameters will be presented. The results of vibration tests to compare frequency, onset pressure, amplitude, and flow rate data of 3D printed models to traditional cast models will then be shown and discussed. Vibration patterns of printed multi-material models characterized using high-speed imaging will also be presented. The results will be used to demonstrate the overall effectiveness and efficiency of embedded 3D printing as an approach for fabricating multi-material VF models. It is anticipated that this process will accelerate the development of multi-material VF models and expand the ability of synthetic VF models to investigate the effects of interior geometry and stiffness on VF vibration.

Acknowledgements

References

[1] Murray and Thomson, J Acoust Soc Am, 132:3428-3438, 2012.
[2] Wehner et al., Nature, 536:451-455, 2016.
[3] Murray and Thomson, J Visualized Exp, 58, 2011.

Multi-axial mechanical properties of hydrogel-based materials upon finite strains: towards the design of tailored vocal-fold composite replicas

H. Yousefi-Mashouf[1,2], D. Ferri-Angulo[3], L. Bailly[1]*, J. Sohier[3], L. Orgéas[1], N. Henrich Bernardoni[2]

[1]Univ. Grenoble Alpes, CNRS, Grenoble INP, 3SR, 38000 Grenoble, France
[2]Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, 38000 Grenoble, France
[3]Univ. Claude Bernard Lyon 1, CNRS, INSA, MATEIS, 69372 Lyon, France

Keywords: vocal-fold replica; tunable cross-linked hydrogel; multi-axial mechanical loading; finite strain

Introduction

In vitro modeling of phonation requires materials able to mimic the vibro-mechanical features of vocal-fold tissues: ability to (i) endure large deformations under physiological multiaxial loadings and strain rates [1], (ii) adapt their vibro-mechanical behavior to external loadings and environmental changes. Among relevant candidates [2], hydrogels are attractive materials due to their tissue-like water content [3,4]. However, the mechanical characterization of hydrogels is often limited to single monotonic loadings, mainly in compression, or to standard DMA analyses [5]. These configurations are far from those endured by the tissue *in vivo*. Here, we manufactured hydrogels with tunable mechanical properties that we characterized under realistic, finite strain, cyclic tension, compression and shear loadings.

Methods

Materials. Two types of hydrogels were used. The first one was based on 10 %w/v porcine Gelatin (Ge) aqueous solution (300 g Bloom, Type A). After molding and freeze-drying, dried plate samples were crosslinked by immersion in Glutaraldehyde (GA) solutions of various concentrations. The second hydrogels were obtained by casting aqueous solutions of poly-lysine dendrimers (DGL) and functionalized polyethylene glycol (PEG-NHS) with various concentrations and molecular weight MW [6].

Mechanical characterization. Hydrogel samples were tested using a tensile testing machine (INSTRON® 5944) equipped with a ±10 N load cell. Three loading modes were studied, as previously done on vocal-fold tissues [7]: simple tension, compression and shear. Samples were first subjected to quasi-static cyclic paths ($10^{-3}s^{-1}$) with increasing strain amplitude up to failure. Shear DMA measurements were also carried out to study the cyclic behavior of hydrogels at various frequencies for small and finite strains. All tests were conducted within a thermo-hygro-regulated chamber at 25°C and saturated humidity (RH ≈100%).

Results and discussion

Typical tensile stress-strain curves recorded for both hydrogels at various compositions are plotted in Figure 1, together with the corridors corresponding to the tensile behavior of the vocal-fold sublayers. Ge-GA hydrogels present proper tensile behavior, of the same orders of magnitude as observed for the *lamina propria*, with weak stress-hysteresis and suitable stiffness up to 10-20% strains. However, their strains at break exhibit weak values which should be improved (up to 30-50%) to approach the mechanics of the *lamina propria*. PEG-NHS hydrogels display similar tensile behaviors, with lower stiffness and higher strains at break, and closer to the behavior of the *vocalis*. By modulating the Ge-GA (resp. PEG/DGL) ratio, the hydrogel stiffnesses can be easily tuned. In addition, increasing the PEG MW allowed the hydrogel elongation at break to be significantly improved, from 30% to 90%.
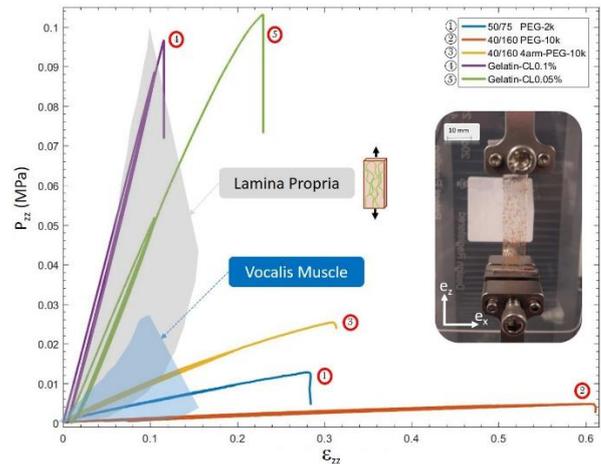
Figure 1: Tensile behavior of the Ge-GA and PEG-NHS hydrogels (Piola-Kirchoff stress P_{zz} vs. Hencky strain ε_{zz}) together with the corridors recorded for the lamina propria and the vocalis [7].

The compression and shear behaviors of the hydrogels are roughly in accordance with those observed for the *vocalis* and the *lamina propria*. Nonetheless, due to their isotropy, hydrogels are stiffer than the highly anisotropic living tissues. modifying the hydrogel nanostructures with proper processing routes or reinforcing them with orientated nanofibers are anticipated solutions to explore.

Acknowledgements

References

[1] T. Vampola et al., *Byocyber Biomed Eng*, 2016, 36(3), 451-465
[2] Gugatschka *et al*, *J Voice*, 8:14003, 2018.
[3] Miri A. *J Voice*, 28, 6:657 2014.
[4] Park *et al*, *Polymers*, 8(1):23, 2016.
[5] Kwon et al, *J Biomech*, 43(3) :420, 2010.
[6] Debret *et al*, *Patent*, WO2017/194761.
[7] Cochereau T., *PhD Thesis*, Univ. Grenoble Alpes, 2019.

Bioengineering and characterization of autologous human laryngeal mucosa transplants

Tanja Grossmann[1]*, Michael Karbiener[1,6], Andrijana Kirsch[1], Ruth Birner-Grünberger[2,3,4] , Barbara Darnhofer[3,4], Luka Brcic[5], Markus Gugatschka[1],

[1]Department of Phoniatrics, ENT University Hospital Graz, Medical University of Graz, Austria
[2]Gottfried Schatz Research Center, Medical University of Graz, Austria
[3]Omics Center Graz, BioTechMed-Graz, Austria
[4]Austrian Center of Industrial Biotechnology, Graz, Austria
[5]Diagnostic and Research Institute of Pathology, Medical University of Graz, Austria
[6]Takeda, Vienna, Austria

Keywords: tissue engineering, vocal folds, 3D cell culture

Introduction

Human vocal fold mucosa possesses an unique multilayered structure and has to withstand extreme mechanical strain enabling undisturbed oscillation [1]. Voice disorders are the most common communication disorders across lifespan where treatment options such as speech therapy as well as phonosurgical methods are still far from satisfying [2]. That is why newer therapeutic approaches focus to restore function on a cellular basis. Laryngeal mucosa transplants are a comprising option for treatment of vocal fold scars.

Methods

Our study comprises to collect healthy human laryngeal and oral mucosa to isolate and cultivate fibroblasts and epithelial cells from either location *in vitro*. These cells are used to bio-engineer constructs under 3D-organotypic conditions that will be evaluated against native laryngeal mucosa by histological analyses and proteomics.
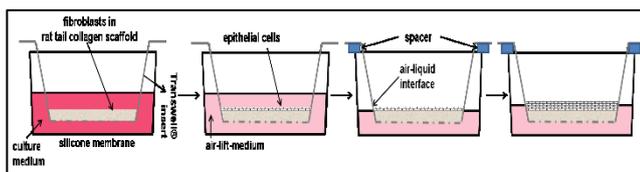
Figure 1: Schematic illustration of the co-cultivation procedure of isolated cells

Acknowledgements

References

[1] Gray SD *et al*, Otolaryngol Clin North Am; 33(4):679–98, 2000
[2] Berry DA *et al*, J Speech Lang Hear Res JSLHR.;44(1):29–37, 2001
[3] Fukahori *et al*, PloS One; 11(1):e0146151, 2016
[4] Karbiener *et al*, Proteomics; 155;11-21, 2017

corresponding author tanja.grossmann@medunigraz.at

Results

Fibroblasts from human vocal folds and epithelial cells from oral mucosa were isolated, cultivated and characterized with q-PCR using cell-type specific marker genes.
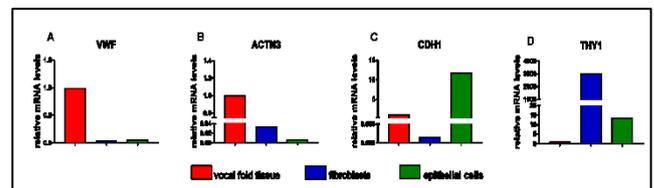
Figure 2. Relative mRNA levels of cell-type specific marker genes of native vocal fold tissue, vocal fold fibroblasts and oral mucosal epithelial cells- Von Willebrand factor (A), actinin alpha 3 (B), cadherin 1 (C) and Thy-1 cell surface antigen (D).

The successful assembly of a multi-layered structure of constructs could be confirmed through histological and immunohistochemical analyses and evaluated against native laryngeal mucosa samples.
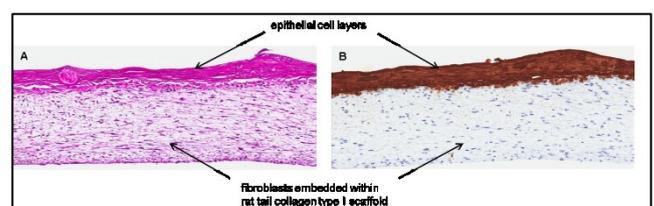
Figure 3: Representative H&E staining (A) and immunohistochemical staining for Cytokeratin 5/6 (B) of formalin-fixed, paraffin-embedded sections of 3D constructs (both 10x magnification)

Discussion

Established isolation procedures of cells will give the opportunity to generate 3D constructs being highly comparable to native vocal fold tissue in histological and proteomic features [3,4]. Besides their implementation as option for treatment of vocal fold scars, constructs will contribute substantially to a deeper understanding of vocal fold micro-physiology and may provide a possible *in vitro* platform for drug testing

Gene expression analysis using RNA sequencing reveals novel insights into pathological alterations in human vocal fold fibroblasts from patients with Reinke's edema

Magdalena Grill[1]*, Isaac Lazzeri[2], Tanja Grossmann[1], Ellen Heitzer[2], Markus Gugatschka[1]

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

Keywords: Reinke's edema; vocal fold fibroblasts; RNAseq

Introduction:

Reinke's edema (RE) is a smoking- and voice-abuse associated benign lesion of the vocal folds, defined by an edema of the Reinke space [1][2][3]. Vocal fold fibroblasts (VFF) are the main cell type of the lamina propria and thought to have a key role in the ongoing of the disease [4][5][6]. In addition, changes in the microvasculature and increased infiltration of immune cells have been described [7]. Current treatment strategies are restricted to treatment of symptoms and there is an urge need for a better understanding of the molecular causes of the disease. In the current study, we investigated differential expression profiles of RE and control VFF by means of RNAseq.

Methods:

Human vocal fold tissue samples from patients with RE and healthy controls were collected during phonosurgical procedures, or received post mortem (Diagnostic and Research Institute of Pathology, Medical University of Graz) (RE n=18, controls n=9). Tissue samples were allowed to adhere to 12-well plates and cultured until outgrowth of fibroblastoid cells. Near-primary VFF were cultured in T25 flasks and RNA isolation was performed using Qiazol and miRNeasy mini kit (Qiagen). Sequencing library was prepared using 500 ng RNA and the TrueSeq® Stranded total RNA HMR kit (Illumina) and sequencing was performed on a Nextseq 550 sequencer. Normalized read counts (TPM) were generated by running Salmon on RNA-seq data. Results from Salmon were used for differential expression analysis by DeSeq2.

Results and Discussion:

We identified 81 differentially regulated genes in total (log2FC > +1 or < -1; adjusted p-value < 0.05), of which 21 genes were upregulated and 60 downregulated.

We subjected the set of differentially regulated genes to gene set enrichment analysis using FGSEA in order to obtain biological processes enriched in VFF of RE patients compared to controls. The top 5 terms for upregulated KEGG pathways revealed lysosome, glycosaminoglycan biosynthesis chondroitin sulfate, glycosaminoglycan degradation, complement and coagulation cascades and ECM receptor interaction.

Gene ontology enrichment analysis of molecular function also showed an increase in extracellular matrix structural constituent, proteoglycan binding, platelet derived growth factor binding, hydrolase activity hydrolyzing o-glycosyl compounds and collagen binding (adjusted p<0.05).

Top 5 relevant downregulated KEGG pathways were DNA replication, oxidative phosphorylation, proteasome, cell cycle and ribosome.

The current study addressed for the first time a direct comparison of RE to control VFF, showing a major increase in pathways associated with extracellular matrix components and a decrease in pathways associated with cell growth and proliferation.

Acknowledgements

References

[1] Zeitels et al. Otolaryngol Head Neck Surg. Apr;126(4):333-348. 2002
[2] Dikkers et al. Ann Otol Rhinol Laryngol 1995 Sep;104(9 Pt 1):698-703.
[3] Marcotullio et al. Am J Otolaryngol 2002 Mar-Apr;23(2):81-84.
[4] Hirano et al. Acta Otolaryngol 1999 Mar;119(2):271-276.
[5] Buckley et al. Trends Immunol 2001 Apr;22(4):199-204.
[6] Martins et al. J Voice 2016 Jul 21.
[7] Sato et al. Ann Otol Rhinol Laryngol 1999 Nov;108(11 Pt 1):1068-1072.

Development and validation of a novel phonomimetic bioreactor

Claus Gerstenberger*, Andrijana Kirsch, David Hortobagyi, Tanja Grossmann, Markus Gugatschka

Division of Phoniatrics, ENT University Hospital Graz, Medical University of Graz, Graz, Austria

vocal fold fibroblasts, bioreactor; cell culture; vibration

Introduction

Vocal fold fibroblasts (VFF) constitute the main cell type of the vocal fold's lamina propria, produce the extracellular matrix and thereby determine the tissue characteristics. To study VFF behavior under *in vitro* conditions it is important to mimic the dynamic environment of the *in vivo* state [1]. The aim of our study was to develop and validate a novel phonomimetic bioreactor system mainly based on commercially available components. The use of cell culture dishes with flexible silicone bottoms in combination with a suitable loudspeaker made it possible to expose the cells to various kinds of phonatory stimuli.

Methods

The vibration system consisted of a loudspeaker mounted into a custom-made polyoxymethylene (POM) housing with cell culture plate fixation elements (Fig 1A). The loudspeaker (AL 170, Visaton GmbH, Haan, Germany) was connected via an XLR audio cable to an audio power amplifier (XLS 1502, 775W, Crown International, Elkhart, IN, USA). The audio cable was pulled through a sealed hole in the back of the incubator (ICO105, Memmert GmbH, Schwabach, Germany). We used 6-well-plates with a flexible silicone base (BioFlex, Flexcell International Corporation, Burlington, NC, USA) which were directly stimulated via sound waves. Stimulation sound files were generated with the open source audio software Audacity (audacityteam.org) and exported to a mobile audio player (iPod touch 4G, Apple Inc., Cupertino, CA, USA). The voltage at the audio output of the audio power amplifier served as a measure of the vibration intensity and was determined with a digital multimeter (Fluke 179, Fluke, Everett, WA, USA). The laboratory setting of the components is shown in Fig 1B. We were able to couple any stimulation pattern with high amplitudes and frequencies up to 10 kHz. An example of a 24 hours pattern is shown in Fig 1C.

The vibration behavior of the silicone membranes in the culture plates was investigated by laser Doppler vibrometry (Polytec OFV 3001, Waldbronn, Germany) [2].

Immortalized human vocal fold fibroblasts (hVFF) [3] seeded on the flexible culture plates and allowed to attach for 24 hours in static conditions and then transferred to the vibration bioreactor for 48 hours. Non-vibrational control cells were cultivated in parallel in a separate incubator. After a one-hour rest period, cells and supernatants were harvested for subsequent analyses.

Results and Discussion

We present a new type of phonomimetic bioreactor. Compared to previous models, our device is easy to assemble and cost-effective, yet can provide a wide spectrum of phonatory stimuli based on the entire dynamic range of the human voice. Depending on the amplifier's set volume and the selected stimulation frequency displacements up to 1 mm in the center of a well could be achieved. Gene expression data of VFF cultured in our phonomimetic bioreactor show a significant effect of vibration on ECM metabolism, which illustrates the efficacy of our device. Gene expression of hyaluronan synthase 2, collagen III, fibronectin and TGFß-1 was significantly upregulated in VFF exposed to vibration, compared to static control. Vibration also significantly upregulated collagen I gene and protein expression.
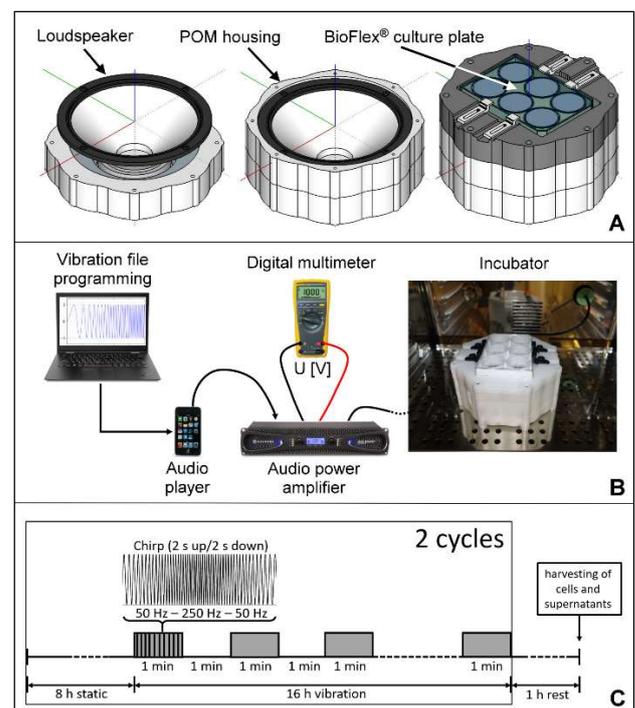
Figure 1: Construction of the bioreactor. A custom-made POM housing, designed using a 3D CAD software, accommodated the BioFlex plate (A). The schematic diagrams depict the assembly of all the components of the bioreactor (B) and the vibration pattern to which the cells were exposed to (C).

References

[1] Catten M *et al*, Otolaryngol Head Neck Surg. 1998 May; 118(5):663–7.

[2] Kirsch *et al*, PLoS One. 2019 Mar 14; 14(3): e0213788

[3] Chen X *et al*, Tissue Eng Part C Methods. 2009 Jun; 15(2):201–12

corresponding author: claus.gerstenberger@medunigraz.at

Session #2 - Flow-induced vibrations: modelling, simulation and validation

Energy-consistent modelling of the fluid-structure interaction in the glottis

Fabrice Silva[1], Thomas Hélie[2], Victor Wetzel[2]

[1] Aix Marseille Univ., CNRS, Centrale Marseille, LMA, Marseille, France
[2] S3AM Team, UMR STMS 9912, IRCAM-CNRS-SU, Paris, France

Keywords : Fluid-structure-sound interactions, voice production modeling

Introduction

The physics of voice has been fruitful in producing a wide variety of models, from full-featured numerical ones (mostly based on FEM for tissues and FVM for airflow) to reduced-order models focusing on the predominant phenomenon. A large body of work in the second category relies on the description of the glottal aerodynamics from the late 1950, based on Bernoulli theorem, thus ignoring rapid or large-amplitude motions of the vocal folds. This is justified on experiments in rigid static larynx-like ducts. However, those simplified models contain an intrinsic paradox where the tissues are driven by the power received by the flow, whose description assumes it does not provide power to tissues.

The main objective of the current communication is to propose a minimal model of the full vocal apparatus that restores a consistent description of the power exchange between the glottal airflow and the vocal folds.

Methods

The model is formulated using the port-Hamiltonian systems (pHS) theory, a framework emphasizing the separation between the intrinsic behavior of components and their interconnection through ports (see Ref. [1] for an introduction). The detailed description of the proposed model is given in Ref. [2]. Its main characteristics is that it includes the simplest kinematics of an incompressible potential flow of inviscid air in the glottis that ensures the continuity of the normal velocity on the surface of the vocal folds (assumed planar for sake of simplicity, see Fig. 1)
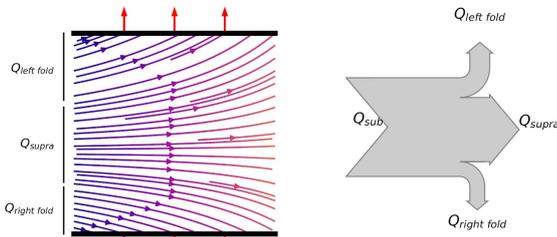
Fig. 1: Example of the velocity field during the glottal opening (on left) and representation of the volume flow division (on right).

The capabilities of the proposed model are explored by means of time-domain simulations (notably using the passive guaranteed numerical scheme derived from the pHS theory) and numerical continuation (giving access to bifurcation and limit cycles, i.e., phonation thresholds and steady-state oscillating regimes).

corresponding author e-mail: silva@lma.cnrs-mrs.fr

Results

Several cases were explored numerically to investigate the effect of the pre-phonatory configuration and of the asymmetry of the vocal folds.
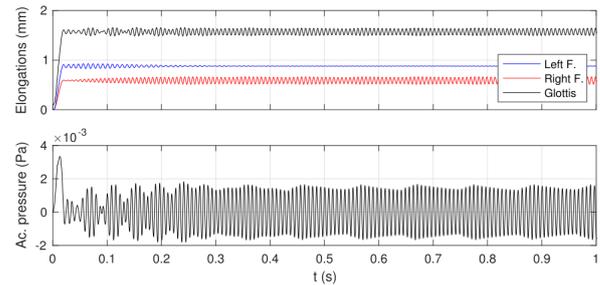
Fig. 2: Example of time-domain simulation in a strongly asymmetric configuration (here, right fold stiffer than left fold)
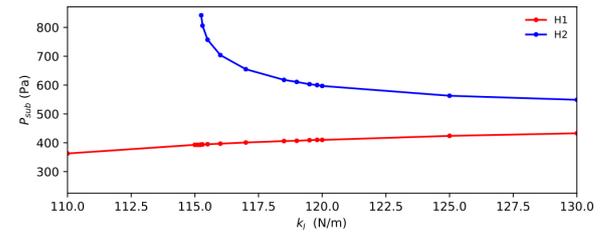
Fig. 3: Bifurcation diagram showing the existence of a second limit cycle beyond some asymmetry threshold

Discussion

Accounting for the transverse component of the glottal flow leads to a more consistent modeling of the fluid-structure interaction that drives phonation. The results obtained with the proposed model already evidence interesting outcomes such as the contact stresses that apply to the superficial layers of tissues during glottal closure, the production of raucous sounds (quasi-periodic regimes) or the existence of a second possible limit cycle for asymmetric folds that can lead to intermittency and thus to pathological voice.

Increasing the complexity of the model will remain easy thanks to the modularity provided by the pHS framework.

Acknowledgements

References

[1] van der Schaft *et al*, Port-Hamiltonian Systems Theory: an Introductory Overview, Now Publishers Inc, 2014.
[2] Hélie *et al*, 3rd conf. Geom. Sci. Info. , 375-383, 2017.
[3] Lopes *et al*, IFAC Papers online 48(13):223-228, 2015.

How the maximum divergence angle of the glottis can affect phonation mechanism

Elias Sundström*, Liran Oren, Charles Farbos de Luzan, and Sid Khosla

Department of Otolaryngology – head and neck surgery, University of Cincinnati, Cincinnati, Ohio, USA

Keywords: phonation mechanism; intraglottal vortices; FSI

Introduction

During vocal fold vibration, there may be a mucosal wave in the superior-inferior (vertical) direction, resulting in a convergent shape during opening and a divergent shape during closing. In our previous work we characterized the dynamics of the medial glottal wall geometry during vibrations in the full excised canine larynx model [1]. We found that the glottal divergence angle during closing is proportional to the magnitude of the acoustic intensity and the intraglottal negative pressure. The latter may act as additional aerodynamic force that contributes to the closing mechanism of the folds' vibration. The magnitude of the negative pressure is affected by the strength of the vortices that develop near the superior aspect of the folds [2]. These intraglottal vortices develop when the flow separates from the glottal wall due to the divergent shape of the glottis.

The goal of the current study is to systematically examine how the maximum divergence angle of the glottis affects intraglottal vortical strength during closing. Our hypothesis is that the vortical strength would be maximized at some critical angle of the divergent glottis. This hypothesis is based on the rationale that vortical strength also depends on the flow proximity to its confined walls. The shear layer of a vortex near a confined wall is smaller compared with a vortex in quiescent flow, which makes the vortex more energetic.

Methods

The geometry for the model extended from the pharynx to the subglottal region. The shape of the vocal folds was based on the M5 model [3]. The false folds were excluded from the study. An idealized vocal fold was considered between the pharynx and the subglottal region. The Navier-Stokes equations were solved numerically for prediction of the flow behavior. Fluid-structure interaction between the vocal fold and the viscous fluid was controlled with a prescribed motion, where the maximum opening occurred at the inferior edge. The motion in the superior aspect was systematically varied so that the divergence angle of the glottis during closing varied from 0 to 100 deg. Local control surface deformation of the vocal fold was handled with mesh morphing [4] and a large relative motion was handled with an overlapping mesh approach to enable full closure.

Results

In-plane vorticity for three different intraglottal divergence angles is shown in Figure 1. For a maximum divergence angle of 0 deg the folds are parallel and the intraglottal flow remains attached to the wall. There is no formation of intraglottal vortices. As the maximum divergent angle increases, the intraglottal vortices begin to form near the superior edge. The maximum vortical strength is observed for a maximum divergence angle of 30 deg. As the maximum divergence angle is further increased, the size of the vortex increases as the shear-layer's detachment point migrates towards the inferior edge. The vortical strength, however, begins to decrease.
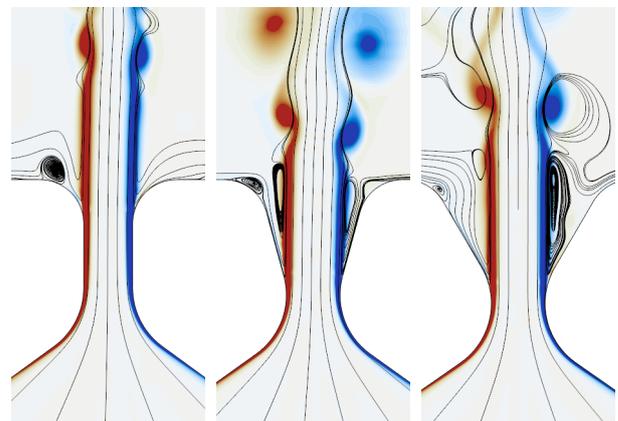
Figure 1: In-plane vorticity with overlayed streamlines in the glottal region. Maximum Intraglottal angles: 0º (left), 30º (middle), and 60º (right).

Discussion

A divergent shape during closing can be clinically achieved by making the glottis stiffer inferiorly. The current study shows that a divergence angle around 30 deg would maximize the intraglottal vortical strength, which is important for the mechanism of folds vibrations.

Acknowledgments

References

[1] Oren L., Gutmark E., and Khosla S. Medial surface dynamics as a function of subglottal pressure in a canine larynx model. J of Voice, [In print].

[2] Oren L., Gutmark E., and Khosla S. Intraglottal pressure distribution computed from empirical velocity data in canine larynx, Journal of Biomechanics, 47, pp. 1287-1293, 2014.

[3] Kucinschi BR, Scherer RC, DeWitt KJ, Ng TT. Flow visualization and acoustic consequences of the air moving through a static model of the human larynx. Journal of biomechanical engineering, 128, pp. 380-390, 2006

[4] Sundström et al. Effects of Normal Variation in the Rotational Position of the Aortic Root on Hemodynamics and Tissue Biomechanics of the Thoracic Aorta. Cardio Eng Tech, 2019 (in-print

12

*corresponding author e-mail: sundstes@uc.edu

Development of a deep learning approach based glottal flow model using high-fidelity numerical simulations on universal vocal fold kinematics models

Yang Zhang[1], Xudong Zheng[1], Qian Xue[1]*, Simeon Smith[2], Ingo Titze[2]

[1]Department of Mechanical Engineering, University of Maine, Orono, ME, USA
[2]The National Center for Voice and Speech, University of Utah, Salt Lake City, UT, USA

Keywords: Glottal Flow; Machine Learning; Deep Neural Network; Reduced-order Modeling

Introduction

An accurate prediction of vocal fold vibration and sound source relies on an accurate prediction of intraglottal pressure distribution and glottal flow rate in flow-structure interactions (FSI). Current computer models of glottal flow during voice production is based on either Navier-Stokes equation, which can provide accurate and detailed flow information but requires extremely high computational cost, or Bernoulli equation, which is simple and fast but has low accuracy. In this research, we aim to develop a deep-learning based glottal flow model that can provide fast and accurate predictions of the intraglottal pressure distribution and flow rate in any glottal shapes during vocal fold vibration. The flow model will be further coupled with a continuum vocal fold model to realize nearly real-time FSI simulations with high accuracy.

Methods

The underlying assumption of the approach is that the vocal fold kinematics can be approximated by a few vibration modes as described by the surface –wave approach in [1]. For normal phonation, the vocal fold vibration is dominated by two modes. The displacement of the glottis over time can be represented by a linear combination of the modal displacement of these two modes as follows:

$$\xi(y, z, t) = \xi_m \sin\left(\frac{\pi z}{L}\right) \left[sin\omega t cos\frac{\omega(y - y_m)}{c} - (1 - \alpha)cos\omega t sin\frac{\omega(y - y_m)}{c} \right] \quad (1)$$

where $\xi_m, \alpha, \beta, \phi$ are the governing parameters to be determined for each shape. A generalized glottal shape library is built through systematic variations of these governing parameters in broad ranges. For each shape in the library, the groundtruth values of the flow rate and pressure distribution are obtained by solving the Navier-Stokes equation. Then, the mapping relationship between the governing parameters (input features) and the corresponding groundtruth values of the flow rate and pressure distribution (output targets) can be established by a fully connected deep neural network (DNN). With this trained DNN, the flow rate and pressure distribution along any glottis shape generated by Eq. (1) can be predicted.

Results and Discussion

To verify that Eq. (1) can be used as a generalized equation to represent any glottal shape during normal phonation, FSI simulations of a three-layer vocal fold model under different subglottal pressures and material properties are conducted. Eq (1) is then used to fit the obtained glottal shapes. The results show a very good agreement. *Figure 1(a)* shows the probability density function (PDF) of the fitted governing parameters. The DNN flow model is then used to predict the pressure distribution and flow rate in these glottal shapes. The prediction is compared with the truth values obtained from the Navier-Stokes flow solver. The preliminary results show good prediction accuracy. *Figure 1(b)* shows the comparison of the pressure distribution in two typical convergent and divergent glottal shapes.

The coupling of the DNN flow model with a continuum vocal fold model is currently untaken. The accuracy of this FSI model in predicting vocal fold vibrations and glottal flow waveform will be compared with Navier-Stokes equation based high-fidelity FSI simulations.

(a)   (b)
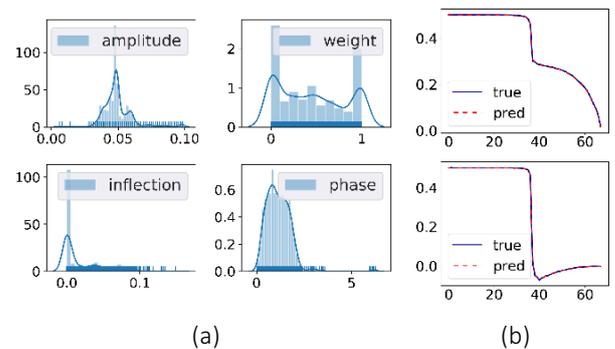
Figure 1: (a) PDF of fitted governing parameters; (b) Performance of DNN pressure prediction

Acknowledgements

References

[1] Simeon Smith and Ingo Titze. Journal of Biomechanics, 73:177–184,2018.

Effect of the paraglottic space on phonation in a MRI-based vocal fold model

Liang Wu, Zhaoyan Zhang*

Department of Head and Neck Surgery, UCLA School of Medicine, Los Angeles, CA, USA

Keywords: Paraglottic space; boundary condition; MRI-based model

Introduction

The paraglottic space refers to the connective tissue layer that connects the vocal folds and the thyroid cartilage. This region is often neglected in computational models of phonation, in which a fixed boundary condition is often imposed on the lateral surface of the vocal fold, with the implicit assumption that the lateral surface of the vocal fold is in direct contact with the thyroid cartilage. Recent studies suggested that the presence of the paraglottic space may have important impact on voice production. The goal of this study was to investigate the effect of the paraglottic space on voice production in a MRI-based vocal fold model, and how this effect may be affected by vocal fold stiffening due to laryngeal muscle activation.

Methods

A parameter vocal fold model was developed based on MRI scanning of a cadaver hemi-larynx (57-year-old male) as described in [1]. The thyroid cartilage was directly imported from the MRI reconstructed geometry, and the paraglottic space was added by filling the gap between the vocal fold and the thyroid cartilage, as shown in Figure 1. The paraglottic space was modeled as an isotropic linear elastic material with varying Young's modulus. The vocal fold was modeled as a transversely isotropic linear elastic material as in our previous studies, with different vocal fold stiffness conditions simulating effects of laryngeal muscle activation. For each vocal fold and paraglottic space stiffness condition, voice production simulations were performed for different subglottal pressure and medial surface shape, from which the phonation threshold pressure, vocal fold vibration amplitude, and the closed quotient of vocal fold vibration were calculated.

Results

The results showed that the presence of the paraglottic space increased the mean and amplitude of the glottal area waveform, decreased the phonation frequency and closed quotient. The effect on the phonation threshold pressure was generally small. In this particular geometry with fully approximated vocal folds, the presence of the paraglottic space also reduced the occurrences of irregular vocal fold vibration. These effects of the paraglottic space generally became smaller with increasing vocal fold stiffening.
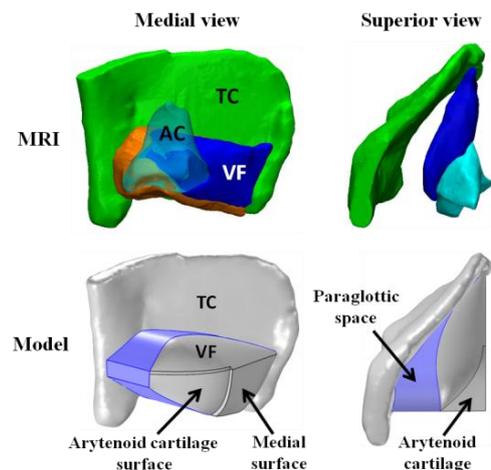
Figure 1: MRI-based vocal fold model with paraglottic space. VF: vocal fold; TC: thyroid cartilage; AC: arytenoid cartilage.

Discussion

This study showed that the stiffness condition of the paraglottic space can have important effect on vocal fold vibration and voice acoustics, and thus should receive more attention in vocal fold computational modeling, particularly in models based on realistic laryngeal geometry.

Acknowledgements

References

[1] Wu, L., and Zhang, Z., J. Acoust. Soc. Am. 140: EL159-165, 2016.

corresponding author e-mail: zyzhang@ucla.edu

Numerical investigation of the integrated neuromuscular control and flow-structure interaction during phonation using an image based high-fidelity computer model

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

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

Keywords: neuromuscular control, flow-structure interaction, image-based modeling

Introduction

Vocal fold vibration is considered largely a mechanical process. The versatility of human voice comes with the ability to change the mechanical status (geometry and internal stress) of the vocal folds through activation of laryngeal muscles. Despite being of high interest to researchers, the relationship between the muscle activation, the mechanical status of the vocal folds, and the resulted flow-structure interactions during phonation is not well understood. Previous simulation works often used simplified vocal fold geometry and eigen analysis [1,2]. The current study aims to use a high-fidelity continuum computer model that couples the muscular mechanics and flow-structure interactions to study the relationship between laryngeal muscle activation, vocal fold geometry and internal stress and the resulted vocal fold vibration pattern in an image-based realistic laryngeal model. The preliminary results have revealed a complex effect of thyroarytenoid muscle activation on the vibration pattern and flow rate.

Methods

The vocal fold model was reconstructed from a MRI scan of a dissected canine larynx shown in Figure 1. For the muscles, both the passive and active stresses were considered. The passive stress was modeled using fiber-reinforced materials and the active stress was modeled using Hill-based contractile elements. Details of the muscle model was presented in a previous work [3]. The muscle model was integrated in a finite element vocal fold model, which was coupled with both Navier-Stokes and Bernoulli flow solvers to simulate the sustained vibration of the vocal folds. Various activation conditions of the intrinsic laryngeal muscles and their effects on flow-structure interactions were investigated.
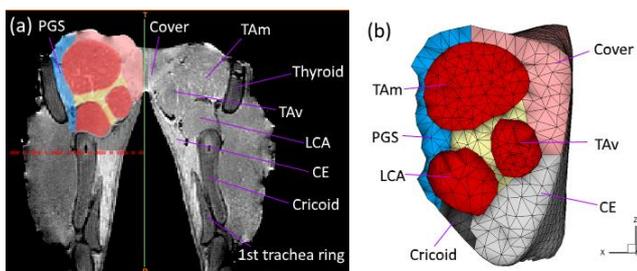
Figure 1: The MRI scan of the canine larynx (a) *and the vocal fold model used in the current study* (b), *only the left side vocal fold is shown.*

Results and Discussion

The glottal waveform and vocal fold vibration pattern from two activation conditions are compared here. In the first condition, the cricothyroid (CT) muscle was activated by 20%. In the second condition, the vocalis muscle (TAv) was activated by 80% on top of the 20% CT activation. Activation of the vocalis increased the fundamental frequency from 144Hz to 162Hz. Frequency increase due to TA activation was also reported in a previous experiment [4]. The maximum flowrate decreased from ~400 mL/s to ~240 mL/s. The results showed that the decrease in the flow rate is caused by the medial bulging at the inferior aspect due to TAv activation which has affected the flow-structure interactions.
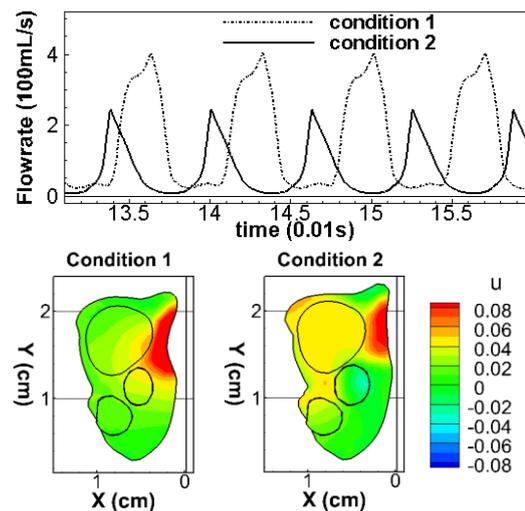
Figure 2: Glottal waveform (top) and mid-coronal profile of the vocal fold at max opening (bottom) for the 2 activation conditions. Contour based on lateral displacement.

The current approach shows potential in revealing the relationship between laryngeal muscle activation and vocal fold dynamics. More activation conditions are currently being investigated.

Acknowledgements

References

[1] Yin *et al*, J Acoust Soc Am, 133(5): 2972-2983, 2013.
[2] Yin *et al*, J Acoust Soc Am, 140(3): EL280-EL284, 2016.
[3] Pham *et al*, J Acoust Soc Am, 144(3): EL248-EL253, 2018.
[4] Chhetri *et al*, J Acoust Soc Am, 135(4): 2052-2064, 2014.

corresponding author e-mail: xudong.zheng@maine.edu

A study on the role of muscle tonicity on the onset of self-excited oscillations in tracheoesophageal speech

André M. C. Tourinho[1]*, Camila Z. M. Araujo[1], Andrey R. da Silva[1]

[1]Department of Mechanical Engineering, Federal University of Santa Catarina, Florianópolis, SC, Brazil.

Keywords: Pharyngoesophageal segment; tracheoesophageal speech; linear stability analysis; collapsible channel.

Introduction

Laryngectomees with either a hypertonic or a hypotonic pharyngoesophageal segment (PES) have difficulties in producing the tracheoesophageal voice. In this work, we investigate how the tonicity of the PES affects the onset of self-excited PES oscillations.

Methods

The PES is modeled as a collapsible channel (Figure 1).
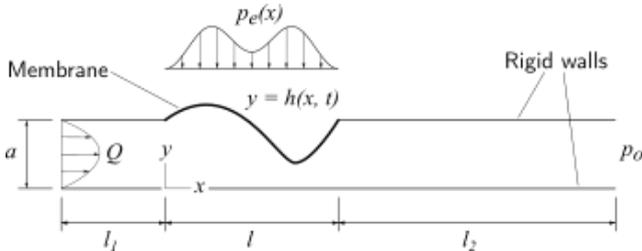
Figure 1: Schematic representation of a collapsible channel.

The simplified formulation of Stewart [1], with few modifications, is adopted. It consists of two dimensionless partial differential equations in x and t. The parameters of the model were set based on published data regarding tracheoesophageal speech. Of special note is the case of the external pressure applied on the membrane (Figure 1), which models the tonicity of the PES. It is set as

$$P_e(x) = \frac{4\bar{P}_e}{3}\left(\frac{1}{2}\sin^2 \pi x + \sin^2 2\pi x\right), \qquad (1)$$

where $\bar{P}_e$ is the mean value of the pressure distribution. This function was chosen as an approximation of the intraluminal pressure in the PES obtained by Welch [2] for laryngectomees at rest. In order to study the onset of self-excited oscillations, first steady-state solutions are obtained, then the governing equations are linearized around the steady-state solution, resulting in an eigenvalue problem. Eigenvalues with a positive real part and a nonzero imaginary part indicate an oscillatory instability. To obtain the steady-state solution, the time derivatives are set to zero in the governing equations, resulting in an ordinary differential equation in x, which was solved using a Chebyshev collocation method [3]. The linearized equations were spatially discretized by the same method, resulting in a generalized eigenvalue problem. The parameter $\bar{P}_e$ is varied, and for each combination of parameters, the stability of the steady-state solution is assessed.

Results and Discussion

The steady-state membrane configuration changes considerably with $\bar{P}_e$. To illustrate, Figure 2 shows the steady-state solutions for $\bar{P}_e = 100$ and $\bar{P}_e = 5000$. These solutions share several similarities with radiographic observations of the PES, of different tonicities, during phonation [4].
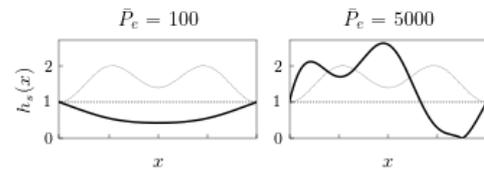
Figure 2: Steady-state solutions for $\bar{P}_e = 100$ and $\bar{P}_e = 5000$. The dashed line shows the form of $P_e(x)$ (not to scale).

Figure 3 shows the eigenvalues of three modes in the complex plane for different $\bar{P}_e$.
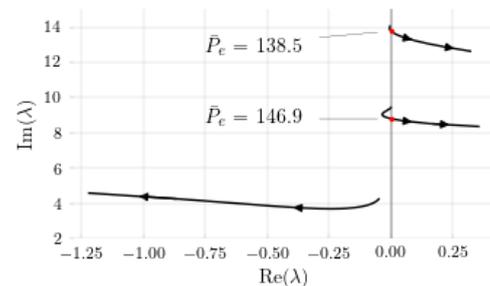
Figure 3: Eigenvalues in the complex plane. The arrows indicate the direction of increasing $\bar{P}_e$.

The onset of self-oscillation happens at $\bar{P}_e = 138.5$. In dimensional units, the corresponding external pressure would be 165 Pa. The frequency for the neutrally stable solution would be 43.5 Hz. While the low threshold may seem to contradict the notion of hypotonicity hindering phonation, there are reports of intraluminal pressures as low as 227 Pa for speaking laryngectomees [5]. Further work is planned to assess the applicability of the model.

Acknowledgements

References

[1] Stewart, P. S., J. Fluid Mech., 825:922-960, 2017.
[2] Welch, R. W. et al., J Clin Invest, 63(5):1036-1041, 1979.
[3] Mai-Duy, N., Comm Numer Meth Eng, 22(6):627-642, 2006.
[4] McIvor, J. et al., Clinical Radiology, 41:312-316, 1990.
[5] Takeshita, T. K. et al., Head Neck, 35(4):500-504, 2012.

*andre.miazaki@posgrad.ufsc.br

Session #3 - Singing

Diaphragm and rip cage movements during phonation of professional singers of different genres – a dynamic MRI study

*Louisa Traser[1,2], Stefanie Rummel[3], Carmen Schwab[4], Ali Caglar Özen[2,5], Michael Bock[2,5], Matthias Echternach[6], Bernhard Richter[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 Prosthetic Dentistry, Center for Dental Medicine, Medical Center – University of Freiburg, Faculty of Medicine, Germany

[5]Deptartment of Radiology, Medical Physics, Medical Center – University of Freiburg, Germany

[6]Division of Phoniatrics and Pediatric Audiology, Department of Otorhinolaryngology, Munich University Hospital, Germany

Introduction

Differences in the respiratory strategy during singing phonation in different musical styles (classic, musical theatre, contemporary commercial music) are postulated in voice pedagogy. Additionally also adaptions of the respiratory system for a specific desired voice quality (e.g. Estill Voice Training® differs 6 voice qualities (speech, falsetto, sob, twang, opera and belting) is suggested [1]. Still scientific examinations in this field are rare and due to technical limitations restricted to indirect or external measures of the breathing system. Latest advances in dynamic Magnetic Resonance Imaging (MRI) enable the evaluation of respiratory movements of diaphragm (DPH) and rip cage (RC) during phonation. First studies using this technique emphasize differences in the movement of the respiratory apparatus between exhaling and phonation [2], as well as task and sex related differences in phonatory breathing movements of professionally trained classical singers [3]. It is also of interest to understand whether and how the movement of the RC and DPH during sustained phonation differs according to the genre of the singer respectively the different voice qualities differentiated by Estill Voice Training®.

Methods

For this purpose 8 professional female singers were included in this pilot study (5 western style classically trained singers and 3 contemporary music singers, trained according to Estill Voice Training®). They were evaluated using a 1.5 and 3 T MRI systems concerning their inner respiratory movements during sustained phonation according to [1]. Classically trained singers were asked to phonate in their stage voice. The Estill trained singers were asked to phonate in 6 different voice qualities according to Estill voice Training® (speech, falsetto, sob, twang, opera and belting). Measurements were acquired in supine position. In a dynamic series of cross-sectional images of the thorax in sagittal orientation characteristic anatomical landmarks were measured. Simultaneous Electroglottography was applied.

Results

Preliminary results show intra-individual differences of the respiratory movements between sustained phonation of different voice qualities, performed by contemporary music singers, trained according to Estill Voice Training®: Here especially belting differed predominant from the other qualities showing a more pronounces rip-cage movement that started at lower lung volumes compared to all other voice qualities. Group differences concerning the respiratory movement of diaphragm and rip cage during sustained phonation for all other voice qualities were minor between contemporary and classically trained singers.

Discussion

A greater understanding of the inner movements of the respiratory system in different genres and for different voice qualities could help to prevent dysbalances in voice production which could lead to vocal overuse. Here greater groups of subjects are desirable for future investigations.

References

[1]   Estill J et al. (1990). Proceedings of the 1990 International Conference on Spoken Language Processing, Kobe, Japan.
[2]   Traser *et al.*, Respir. Physiol. Neurobiol., 236: 69–77, 2017.
[3]   Traser er al., PEVOC conference proceedings 2019.

*louisa.traser@uniklinik-freiburg.de

*louisa.traser@uniklinik-freiburg.de

Performance efficiency evaluation on professional artists of Baroque Repertoire through simultaneous monitoring of vocal doses and breathing pattern

Francesca Cunsolo[1]*, Valeria Ottaviani[1], Silvia Capobianco[2], Orietta Calcinoni[3], Raffaele L. Dellacà[1]

[1] Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milan, Italy
[2] Facoltà di Medicina e Chirurgia, Università degli Studi di Pavia, Pavia, Italy
[3] Voice & Music Professionals Care Team – VMPCT C/o Medico Turati, Milan, Italy

Keywords: vocal dosimetry; singing voice; respiratory kinematics; phonation

Introduction

The voice production involves the synergic cooperation of the respiratory and the vocal systems. Among voice professionals, singers need to focus on efficient phonation strategies to optimize voice production in their artistic exhibitions [1]. A wearable tool providing quantitative evaluation of breathing pattern and voice production would be very helpful, as current quantitative assessment approaches are based only on the analysis of vocal parameters and do not consider the efficiency of respiratory muscles in producing the sounds. The aim of this study is to develop a wearable device for the simultaneous measurement of vocal doses and breathing pattern in order to evaluate the interaction between the respiratory and vocal systems in singing voice production.

Methods

The device included an electronic circuit integrating a skin accelerometer for measuring vocal parameters together with an abdominal (AB) and rib cage (RC) respiratory inductive plethysmography belts for monitoring breathing without interfering with singing (Figure 1).
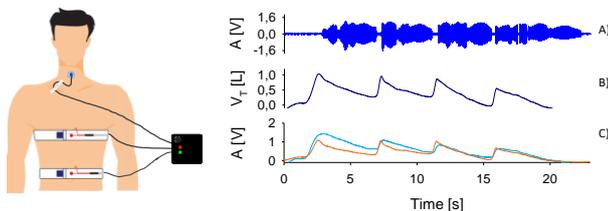
Figure 1: wearable device (left panel) and a representative tracing showing the amplitude of skin acceleration (A), total respiratory volume (B) and its subdivision in RC and AB components (C).

Nineteen professional singers of Baroque repertoire were enrolled in the study together with 18 control (not trained to sing) subjects. All participants were studied during the execution of a series of vocal tasks and of the Italian air "Caro Mio Ben" [2]. The computation of vocal doses [3] and the characterization of the expiratory phase (as in Figure 2.A) were used to evaluate the performance of both groups. The study was approved by the Ethical Review Board of the Politecnico di Milano.

Results

Compared to controls, in singers the RC provided a lower contribution to the overall volume displaced at every phase of the expiration (Figure 2.B).
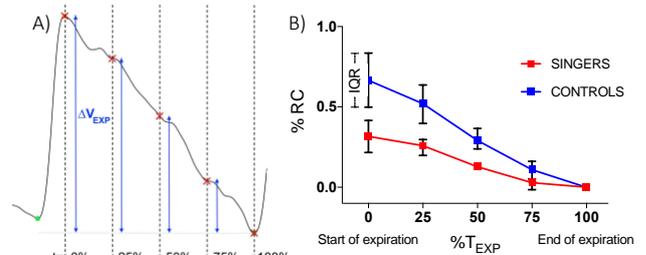
Figure 2: Reduction of the total respiratory volume with time during a singing phrase (A) and percentage of RC contribution to total volume averaged in singers and controls (B).

The RC contribution for exhaling the first 25% of the expiration is compared with Distance Dose (D_D) and the Energy Dissipation Dose (D_E) showing that the effective amount and quality of voice produced is higher in the group of singers (Figure 3).
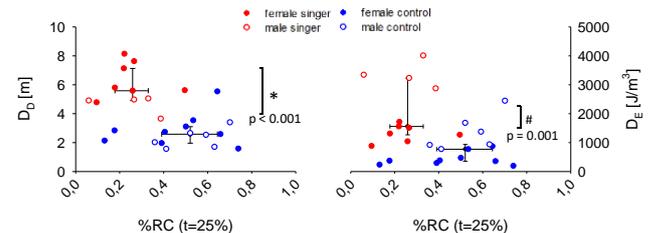
*Figure 3: Comparison between vocal doses (D_D and D_E) and RC contribution at 25% of expiration in singers and control subjects (p-value of t-test * or Mann-Whitney Rank Sum Test #).*

Discussion

Our data suggest that artists reduce the use of the RC in favor of abdominal compartment which, being more compliant, provides airflow with lower effort. The region of the graphs in Figure 3, where most of the singers' data points are positioned, may identify an optimal breathing/singing strategy. This approach may support future research for developing a better understanding of the synergistic interactions between the respiratory and vocal systems, as well as for providing a useful training tool for evaluating the efficiency of the performances.

Acknowledgements

References

[1] J. Sundberg, NATS J, 4-9;100-107, 1993
[2] J. Ralston, JRME, 47(2), 163-173, 1999
[3] I.R.Titze *et al*, JSLHR, 46:919-932, 2003

francesca.cunsolo@mail.polimi.it

Inverse Vocal Tract Adjustment: Spectral Dependence and MRI Data

Patrick Hoyer [1]*, Simone Graf[2], Seiji Adachi[3], Michael Gruner[4], Manuel Graf[5], Louisa Traser[6]

[1]Fraunhofer Headquarters, Munich, Germany
[2]Department of Otorhinolaryngology, Klinikum rechts der Isar, Technische Universität München, Germany
[3]Fraunhofer Institute for Building Physics, Stuttgart, Germany
[4]Fraunhofer Institute for Material and Beam Technology, AZOM Zwickau, Germany
[5]Department of Experimental Physics 5, University of Würzburg, Würzburg, Germany
[6]Institute of Musicians' Medicine, University Medical Center – University of Freiburg, Faculty of Medicine, Germany

Keywords: Inverse Source Filter Model; Formant Tuning; 3D-MRI; Simulation

Introduction

Voice production is a complex process including respiratory adaptions, vocal fold vibration and shaping of the vocal tract. Recently, we proposed a novel route to the specific adjustment of the vocal tract without phonation [1]. The approach is based on an inversion of the source-filter-model were the sound source is placed outside the open mouth and the vocal tract adjusted in order to amplify the resulting sound without phonation. As the vocal tract is essential for resonance and timbre, the technique may enable a training toward defined spectral energy distributions.

Methods

One male subject adjusted the vocal tract toward high amplification of up to three distinguished sinusoidal frequencies related to vowels [a] and [e] in German pronunciation. MRI Data were obtained according to [2]. The acoustic data were recorded simultaneously using a non-magnetic microphone. The procedure of the simulation is described elsewhere [3].

Theoretical Background, Results and Discussion

During the amplification of the sinusoidal frequencies, the detected phase/intensity relation can be explained within the concept of forced oscillation with the speaker being the driving source and the vocal tract being the oscillator [4]. Following this basic model, an amplification is possible if (i) the frequency of the acoustic signal matches the resonance frequency of the resonator (vocal tract) and (ii) the damping of the vocal tract is low. A training towards higher amplification of externally offered sinusoidal frequencies at resonance conditions should thus result in lower acoustic losses. Due to the frequency dependence of the vocal tract impedance, training at selected frequencies should result in different spectral characteristics if the shape of the vocal tract stays constant during subsequent phonation as well.

Fig. 1 shows MRI images of the vocal tract towards adjusted to different sinusoidal frequencies related to the vowel [e] with normal breathing and without phonation.

*Figure 1: MRI cross-sections after adjustment to an external sinusoidal frequency (from left to right): 350 Hz; 350 and 1850 Hz; 1850 and 2700 Hz **without phonation and open glottis.***

Fig. 2 shows the MRI obtained directly after the adjustment shown in Fig. 1 but during phonation.

*Figure 2: MRI cross-sections after adjustment to an external sinusoidal frequency (from left to right): 350 Hz; 350 and 1850 Hz; 1850 and 2700 Hz **during phonation.***

Changes related to the adjustment frequencies occur at full length of the vocal tract. Comparing the images without and with phonation (Fig.1 and 2 respectively) changes are observed mainly at the supralaryngeal tube area, which is more narrow during phonation. In contrast, tongue, jaw and velum position remain approximately stable. An adjustment at higher frequencies results in a shifted tongue position towards the front corresponding to a modified vocal output in subsequent singing.

In the presentation, we offer an acoustical analysis from recorded voices, the corresponding area function and simulation results.

References

[1] Hoyer *et al*, Journal of Voice, 33: 482–489, 2019
[2] Traser *et al*, JSLHR, 33: 2379-2393, 2017
[3] Takemoto *et al*, J. Acoust. Soc. Am. 134: 2955-2964, 2013.
[4] Graf *et al*, Journal of Voice in press (j.jvoice.2018.01.018)

Electroglottographic assessment of Tahrir, a persian vocal technique

Michèle Castellengo[1]*, Jean During[2], Nathalie Henrich Bernardoni[3]

[1]Sorbonne Université, CNRS , Institut Jean-le-Rond d'Alembert, équipe LAM, Paris, France
[2]Univ. Paris-Ouest Nanterre, LESC-crem, UMR -7186 CNRS, 92023, Nanterre, France
[3]Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, F-38000 Grenoble, France

Keywords : Iranian tahrir; laryngeal mechanism; glottal contact; tekye

Introduction

One of the most remarkable characteristics of Persian classical singing is the *tahrir* technique, an ornament of the melodic line which consists in producing one or more short frequency jumps - called *tekye* - towards higher pitches. The change in laryngeal mechanism underlying *tekye* has been reported by different authors [1,2,3-4]. It has recently been questioned [5], yet without laryngeal-behaviour assesment. To clarify laryngeal physiology in Iranian tahrir, the singing production of two professional Iranian singers was acoustically and electrographically assessed.

Method

The singers were Master M.-R. Shajarian, who is the greatest Iranian singer of his time, and one of his disciple Mrs Solmaz Badri. Data recordings were made in laboratory (Paris). Each singer was asked to produce pedagogical vocal exercises, classical ornamental sequences, and freely-chosen musical excerpts. Audio signal and EGG signal were simultaneously recorded (microphone B&K, Nexus conditioning amplifier, EG2 Glottal Enterprise). Several parameters were measured to characterize the singer's laryngeal mechanism [6,7]: fundamental frequency (f0), EGG and differentiated (DEGG) amplitude, EGG-derived contact quotient (CQ).
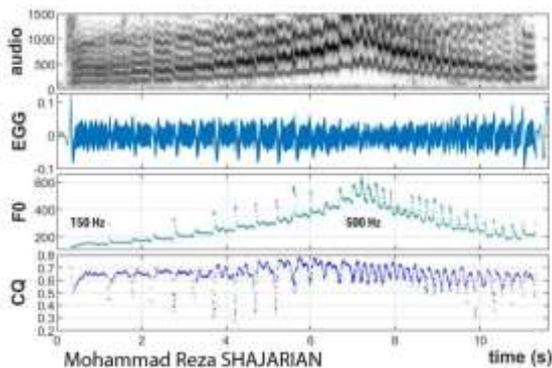
Figure 1: Ascending-descending scale (D3#-B4) with tahrir.

Results

For both singers, *tekye* realization goes with a sudden decrease in glottal contact (EGG amplitude; glottal contact), as illustrated in Figures 1 and 2. It highlights a change in laryngeal vibratory mechanism [7,8].

Main melody is produced in M1 – CQ close to 0,5 or higher; *tekye* jump is a short excursion in M2 – CQ always lower than 0,5. In the ascending part, transition is abrupt (2 or 3 periods). It is less abrupt in descending part. *Tekye* is produced on a great frequency range, from 150 to 500 Hz. Jump duration depends on its definition [5]. The sound part in M2 varies from 25 ms to 80 ms depending on the singing style.
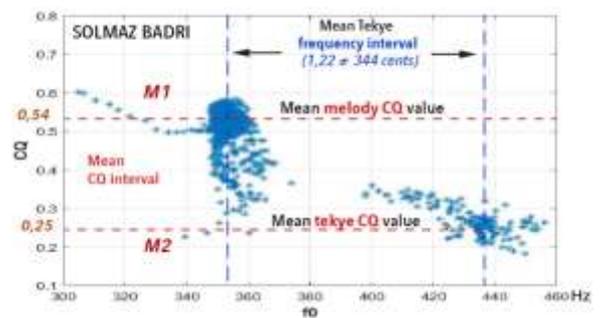
Figure 2: Production of 15 tekye on the same note (F4).

The frequency-jump musical interval is around a third for S. Badri (Figure 2). For M.-R. Shajarian, *tekye* frequency jump can vary from a minor third (315 cents) to a major seventh (1000 cents), with a flexible and changing melodic form.

Discussion

EGG and glottal-contact measurements confirm our previous observations on laryngeal-mechanism jumps in tahrir [4]. The switch between M1 and M2 can be done on ranges as large as two octaves, attesting to exceptional laryngeal control. The two singers adjust *tekye* parameters (duration, melodic shape, interval) to musical context. The diversity of *tekye* melodic features found in M.-R. Shajarian testifies to wide expressive possibilities of this famous artist.

Acknowledgements

References

[1] Caton M., Selected reports in Ethnomusicology, 2(1):42-52, UCLA, 1974.
[2] During J., Musiques d'Iran:347p, Geuthner, Paris, 2010
[3] Castellengo M. *et al*, CIM07:53-55, Tallin, 2007.
[4] Castellengo M. *et al,* PEVOC09:Poster, Dresden, 2009.
[5] Tahamtan M. *et al,* J Voice, in print, 2019.
[6] Henrich N. *et al*, JASA, 117(3):1321-1332, 2004.
[7] Roubeau B. *et al*, J Voice, 23(4):425-438, 2009.
[8] Henrich N. *et al*, JASA, 117(3):1417-1430, 2005.

How voice production of singers is influenced by room acoustics

Paul Luizard[1]*, Silvain Gerber[2], Nathalie Henrich Bernardoni[2]

[1]Audio Communication Group, Technische Universität Berlin, Einsteinufer 17c, Berlin, D-10587, Germany
[2]Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, 38000 Grenoble, France

Keywords: singing voice, singer adaptation, electroglottography, room acoustics

Introduction

Performers, whether instrumentalists or singers, tend to develop rather individual adaptation patterns across room acoustical conditions [1,2]. While instrumentalists mostly adapt in terms of tempo [1], singers tend to emphasize variations of loudness and timbral colour in conjunction with related room-acoustical parameters, namely Stage Support and Bass Ratio [2]. Adaptation to room acoustics has also been investigated on talkers in terms of vocal effort, explicitly mentioned by participants and correlated to voice intensity [3]. Vocal effort while talking was found to be larger in rooms that lack early reflections.

Such adaptation should be reflected on glottal-source parameters. Electroglottographic measurements enable to non-invasively assess several features of glottal behaviour, such as contact quotient (duration of glottal contact over a cycle) or its counterpart, open quotient [4]. In speech, contact quotient was found to be correlated to vocal effort [5]. In singing, it was found to be strongly related to vocal intensity in laryngeal mechanism 1, and to fundamental frequency in mechanism 2 [6].

The present study aims at investigating the singing adaptation process in rooms with various acoustics, by assessing voice production by means of acoustical and electroglottographical in-situ measurements.

Methods

Singing-voice features (sound intensity I, fundamental frequency F0, contact quotient CQ) were compared with room acoustical parameters (early decay time EDT, early stage support ST_{early}, bass ratio BR, late interaural cross-correlation $IACC_{late}$, speech transmission index STI). Eight room acoustics were tested to find correlates between singing-voice parameters and room-acoustics ones which would reveal adaptation patterns. Four singers (soprano, mezzo-soprano, tenor, baritone) were asked to sing exercises (glissandi, crescendi) and excerpts of lyrical musical pieces in each room. Both near-field acoustical signal and electroglottography (EGG) were recorded, synchronised, and calibrated.

The statistical analysis of the data was based on the linear-mixed-models (LMM) framework. This approach can take advantage of the data nested structure and account for random factors, such as different singers and different musical pieces.

Results and discussion

Descriptive statistics showed tendencies, such as an influence of ST_{early} variation (sound level of the room response) across rooms on voice-range profile for two singers, as depicted in Fig. 1 for Singer 3. These results were confirmed and generalized by the LMM analysis. Singers adapted particularly well in terms of vocal intensity when performing singing exercises. The musical pieces yielded larger correlation with room acoustics, namely the proportion of variance in voice production that could be explained by room acoustics, on average across singers: 30% for I, 6% for CQ, and 4% for F0. It should be noted that the variance across singers is large, e.g. for I: 70% for Singer 3 and 2% for Singer 2, indicating that they were not sensitive to room acoustics in a similar manner, each singer having an individual adaptation pattern.
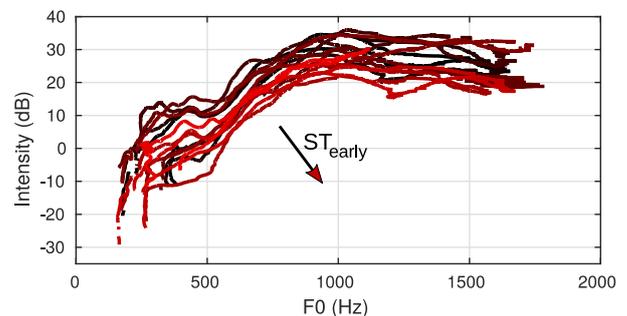
Figure 1: Voice-range profile of Singer 3 for increasing (black to red) room acoustical parameter ST_{early}

Acknowledgements

References

[1] Luizard *et al*, Proc. Conf. AES on Immersive and Interactive Audio, 2019.
[2] Schärer Kalkandjiev *et al*, Psychomusicology: Music, Mind, and Brain, 25(3), 2015.
[3] Bottalico *et al*, J. Acoust. Soc. Am., 139(5), 2870-2879, 2016.
[4] Henrich *et al*, J. Acoust. Soc. Am., 115(3), 1321-1332, 2004.
[5] Huang *et al*, J. Voice, 9(4), 429-438, 1995
[6] Henrich *et al*, J. Acoust. Soc. Am., 117(3), 1417-1430, 2005

**paul.luizard@campus.tu-berlin.de*

Session #4 - Voice analysis and synthesis

GFM-Voc: a tool for analysis and modification of the glottis signal

Olivier Perrotin[1]*, Ian V. McLoughlin[2]

[1]Univ. Grenoble Alpes, CNRS, Grenoble INP, Grenoble, France
[2]University of Science and Technology of China, Hefei, P.R. China

Keywords: glottal flow model; source-filter decomposition; voice quality

Introduction

Estimation of the glottal source signal from recorded voice has been a long-sought goal in voice research [1]. Most recent methods for glottal inverse filtering have addressed the issue of separating the first vocal tract (VT) resonance from the glottal formant, i.e., the low-frequency resonance that describes the open phase of the vocal fold vibration. However, the glottis signal is broadband and few methods are capable of estimating the high-frequency glottis spectral tilt, correlated to the closing phase of the vocal fold vibration, and crucial to vocal force perception [2]. Moreover, such methods often estimate the glottis signal in the time domain [3]. We present here the Glottal Flow Model-based vocoder (GFM-Voc) [4] that allows the real-time extraction and modification of both VT and glottis contributions to the voice signal in the frequency domain, as compact sets of filter parameters. In particular, the glottis filter has a wide-band frequency response, including both glottal formant and spectral tilt characteristics that are closely related to vocal force perception [2]. GFM-Voc thus provides a straightforward way to analyse voice quality, and allows the real-time modification of VT and glottis filters before re-synthesis. Examples of applications include expressive speech synthesis ; auditory feedback perturbation ; and speech therapy.

Methods

A key ingredient of GFM-Voc is the analysis part performed by an improved Iterative Adaptive Inverse Filtering (IAIF) method [5] based on a Glottal Flow Model, which we call GFM-IAIF [6]. Basic IAIF separates a speech frame into 4 components: a flat-spectral-envelope excitation, a high-order linear prediction (LP) VT filter, a low-order LP glottis filter, and a first order high-pass lip radiation filter. In particular, glottis and lip radiation are combined into a first order filter, that reduces the glottis filter to the glottal formant contribution [5]. To include the spectral tilt in the glottis filter, a recent method called Iterative Optimal Pre-emphasis (IOP)-IAIF [7] uses an unconstrained order to encompass all the slope of the speech frame spectrum in the glottis filter. While the IOP-IAIF improvements are merited, we believe that a high filter order risks endowing the glottal model with too much complexity. Our GFM-IAIF method is built on the assumption that a third order filter is enough to model both glottal formant (with a complex conjugate pole pair) and spectral tilt (with one real pole), based on the findings of [2]. A Matlab implementation of

GFM-IAIF is available in [8]. GFM-Voc then implements a full analysis-resynthesis pipeline [4]. GFM-IAIF extracts the glottis and VT filters, that are then modified by changing the position of their poles. Changing the glottis filter poles is equivalent to varying the position and bandwidth of the glottal formant, and position of spectral tilt, resulting in modification of voice quality (vocal force and tension). Changing the VT poles allows the shifting of formants that correlates with changes of jaw, tongue, and lip positions. Finally, the speech signal is reconstructed by filtering the unchanged excitation with the modified filters.

Results

GFM-IAIF has been evaluated against IAIF and IOP-IAIF on both synthetic and natural speech, various vowels and voice qualities (from soft to loud voice), and male and female speakers [6]. Overall, we observed similar performances for estimation of glottal formant parameters. However, GFM-IAIF provided spectral tilt parameters that were closest to ground truth and more discriminative for voice quality than other methods. Specifically, IOP-IAIF and IAIF tend to attribute too much and not enough spectral tilt to the glottis, respectively. Finally, a demonstration of the full GFM-Voc framework for real-time voice modification can be found in [9].

Discussion

GFM-Voc is the first framework allowing high-quality and real-time modification of vocalic formants and voice quality. It relies on GFM-IAIF that extracts vocal tract and glottis contributions as an intuitive set of filter parameters. This method is frequency-domain based as it describes the glottis with spectral parameters, and was evaluated on those. It is not yet investigated how well GFM-IAIF can extract a glottal waveform that is faithful to the temporal vibration of vocal folds, and this is let for future work.

References

[1] Drugman *et al*, Computer Speech & Language, 28(5): 1117-1138, 2014.
[2] Doval *et al*, Acta Acustica, 92(6): 1026-1046, 2006.
[3] Degottex *et al,* Speech Communication, 55: 278-294, 2013.
[4] Perrotin *et al*, Proc. of Interspeech, 3685-3686, 2019.
[5] Alku *et al*, Speech Communication, 11(2-3): 109-118, 1992.
[6] Perrotin *et al*, ICASSP, 7160-7164, 2019.
[7] Mokhtari *et al*, Speech Communication, 104: 24-38,2018.
[8] *http://github.com/operrotin/GFM-IAIF*, (online).
[9] *http://youtu.be/9djf7ljkrsY,* (online).

corresponding author e-mail

Effects of constitutive and dynamic properties of the vibrating vocal folds on vocal frequency perturbations

Jean Schoentgen[1]*, Philipp Aichinger[2]

[1]B.E.A.M.S., Université Libre de Bruxelles, Brussels, Belgium
[2]Department of Otorhinolaryngology, Division of Phoniatrics-Logopedics, Medical University of Vienna, Vienna, Austria

Keywords: vocal jitter; vocal flutter; body-cover model

Introduction

The object of the presentation is the modelling of vocal jitter and flutter in type I voices, which are voices that are monophonic and pseudo-periodic. A formal model of the cause of vocal jitter and flutter has been proposed by Titze [1]. It involves a simulation of the fluctuations of the tension of the TA-muscle (a.k.a. the "body" of the vocal fold) owing to the spatial and temporal superposition of muscle twitches.

We would like to argue that the fluctuations of the tension of the TA-muscle may cause observed vocal frequency perturbations, but that the latter are not a copy of the former. Titze's model may explain, as a consequence, a wider range of phenomena and dispense with the need to postulate the appearance of novel causes of perturbations each time increased jitter or mild hoarseness are observed in type I voices.

The impetus for the investigation is the common experience that vocal loading together with dehydration, smoking, (pre-)menstruation in women or light laryngitis may increase vocal jitter and cause mild hoarseness in type I voices. Fourcin has proposed a physiological model of the former, which involves injecting atropine in the folds [2]. The observation that atropine injection increases vocal jitter whereas, as a rule, moderate vocal loading decreases jitter, suggests that constitutive or dynamic properties of the folds may alter observed perturbations. Fourcin has indeed assigned the boost of jitter to an increase of the damping of the cover.

Method

We consider the predictions of three existing models and report two simulation experiments.

(A) A first model kinematically simulates the vibration of the edge of a vocal fold by the sum of a constant abduction and two sinusoids (body and cover). The phase of the sinusoid mimicking the vibration of the fold body is perturbed feebly.
(B) The second model involves the derivative of a formula in [3] that relates the active and passive tension of the vocal fold to its natural frequency of vibration.
(C) The third model is a lumped three mass model of the vocal folds [4], which involves 9 control parameters the values of which are randomly chosen in physiologically plausible intervals. All signals have been type I exclusively. The purpose of a *first* set of 1000 *perturbation-free* simulations of sustained [a] sounds has been to *interpret* via model (C) the predictions of models (A) and (B). The purpose of a *second* set of 1000 simulations has been to *reproduce* via model (C) the predictions of models (A) and (B) by perturbing randomly the stiffness of the body and recording the perturbations of the frequency of vibration of the body and cover.

Results and discussion

(i) Models (A) and (B) predict that observed frequency perturbations evolve proportionally to the relative amplitude $a_b/(a_b+a_c)$. That is, observed perturbations increase when the body and cover amplitudes of vibration a_b and a_c respectively increase and decrease. In addition, model (B) predicts that observed perturbations decrease with the natural frequency of vibration of the folds.

(ii) Regression analysis of the *first* set of simulations shows that the relative amplitude $a_b/(a_b+a_c)$ evolves proportionally to the relative vibrating mass $m_c/(m_b+m_c)$, which is the most influential control parameter by far.

(iii) Regression analysis of the *second* set of simulations confirms the pre-eminence of the relative vibrating mass. The second-most important parameter that boosts observed frequency perturbations is pulmonary pressure. The regression weights of the remaining parameters are small and they (damping included) attenuate perturbations. The damping of the cover has a major influence on the open quotient and average glottal area, however.

In the framework of model [4], increased damping of the cover therefore favours "breathiness" via the increase of the open quotient, whereas excess secretions on or swelling of the cover favour "roughness". Increased damping may boost "roughness" indirectly provided that the speaker compensates increased "breathiness" by increased pulmonary pressure.

References

[1] I. R. Titze, "A Model for Neurologic Sources of Aperiodicity in Vocal Fold Vibration" J. Speech & Hear. Res., Vol. 34, 460-72, 1991

[2] E. Abberton, A. Fourcin, "Electrolaryngography", in "Instrumental Clinical Phonetics", ed. M. J. Ball, Wiley, 1997

[3] I. R. Titze, Vocal fold mass is not a useful quantity for describing Fo in vocalization, J. Spe. Lang. Hear. Res., 54(2), 520-522, 2011

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

*jschoent@ulb.ac.be

Non-invasive evaluation of vibratory kinematics of phonation in children

Rita R. Patel[1]*, Sten Ternström[2]

[1]Department of Speech & Hearing Sciences, Indiana University, Bloomington, Indiana, USA
[2]Division of Speech, Music, & Hearing, KTH Royal Institute of Technology, Stockholm, Sweden
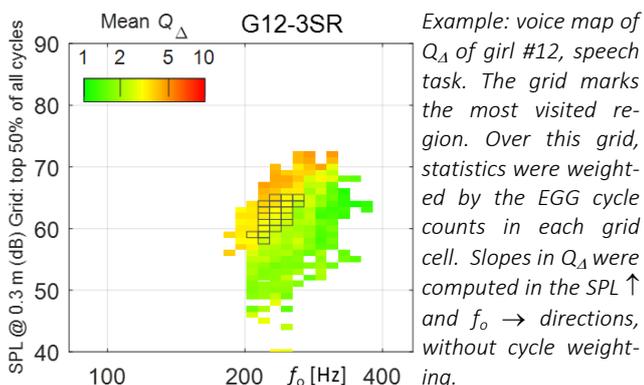
Keywords: electroglottography; voice maps; pediatric voice; FonaDyn

Introduction

Developmentally, children do not have five well-defined layers of the vocal folds as adults. Also, children are often less able to tolerate the invasiveness of endoscopy. Electroglottography (EGG) is not invasive, but little is known of how the EGGs of children differs from those of adults. The goal of this study was to quantify some differences in the shape of the EGG waveform between children and adults, while accounting for the sensitivity to variation in the independent variables f_o and SPL. A novel mapping of EGG waveform parameters over the speech and voice ranges was employed [1].

Method

22 vocally healthy children (9 boys, 13 girls) aged 4-8 years (mean 6.3) were recruited. 26 vocally healthy adults (13 men, 13 women), 22-45 years (mean 28.7) were recruited to obtain developmental end point of maturation of vibratory kinematics. Simultaneous recordings were made of EGG and airborne voice (44.1 kHz/16 bits) at calibrated SPL in dB(C) [2]. Habitual speech production for the speech range profile (SRP) was elicited for adult participants by 3 trials of reading the Rainbow Passage; while child participants were instead shown drawings of standardized outdoor scenes and asked to describe the scenes in detail. Then, phonations on the vowel /a/ were elicited so as to approximate a full voice range profile (VRP) [3]. A custom public-domain analysis software [1] was used to compute the audio crest factor, plus three EGG waveform variables [4]: (1) the quotient of contact by integration (Q_{ci}) (2) the normalized peak dEGG (Q_Δ), and (3) the index of contacting (I_c). In the SRPs, the 2-D region around the mode of the phonation density map which contained 50% of *all* EGG cycles was delineated, see figure. Over this small, 'most typical' region, the means, standard deviations and local slopes of the four dependent variables were computed.

Example: voice map of Q_Δ of girl #12, speech task. The grid marks the most visited region. Over this grid, statistics were weighted by the EGG cycle counts in each grid cell. Slopes in Q_Δ were computed in the SPL ↑ and f_o → directions, without cycle weighting.

Results and Discussion

The table shows partial results from the SRPs only, averaged over subjects within categories. The full/half stars denote significance with/without Bonferroni correction, and **bold** denotes a non-zero gradient ($p<0.05$). There was no difference between girls and boys in any of the metrics. All dependent variable means were slightly lower in children than in adults. All Q_{ci} and I_c slope metrics were positive in the group of men, but negative in the other groups. The crest factor increased with SPL in all groups except women. The results from the VRP, an unfamiliar task over the widest elicitable range, exhibited far more variation over the voice maps, and must be discussed qualitatively, which is out of scope here.

Variable		Unit	women	men	girls	boys	f/m	g/b	g/f	b/m
f_o	mean	Hz	190.2	116.0	253.1	246.5	★		★	★
SPL	mean	dB	71.67	73.58	68.46	69.19			☆	★
Q_{ci}	mean	-	0.430	0.461	0.408	0.405				★
	slope vs f_o	1/ST	-0.86%	0.54%	**-1.78%**	**-3.60%**	☆			★
	slope vs SPL	1/dB	-0.34%	**0.59%**	**-1.04%**	**-1.16%**	★			★
Q_Δ	mean	-	4.28	4.38	3.69	4.06			★	
	slope vs f_o	1/ST	-0.67%	1.59%	-0.78%	-2.34%				☆
	slope vs SPL	1/dB	-0.26%	0.72%	**0.98%**	0.69%			☆	
I_c	mean	-	0.269	0.292	0.226	0.242			★	☆
	slope vs f_o	1/ST	-1.57%	**1.93%**	**-2.67%**	**-5.61%**	★			★
	slope vs SPL	1/dB	-0.63%	**1.25%**	-0.08%	-0.46%	★			☆
Crest	mean	-	2.211	2.947	2.023	2.012	★		★	★
	slope vs f_o	1/ST	**-1.11%**	0.19%	**0.84%**	1.48%			☆	
	slope vs SPL	1/dB	0.00%	**1.70%**	**1.63%**	**1.25%**	★		★	☆

Acknowledgements

References

[1] Ternström S, Johansson D, Selamtzis A. FonaDyn — A system for real-time analysis of the electroglottogram, over the voice range. *SoftwareX* 7: 47-80, 2018

[2] Patel RR, Awan SN, Barkmeier-Kraemer J, Courey M, Deliyski D, et al. Recommended Protocols for Instrumental Assessment of Voice. *Am J Speech Lang Pathol*, 27(3): 1-19, 2018

[3] Pabon, P. *Mapping Individual Voice Quality over the Voice Range, The Measurement Paradigm of the Voice Range Profile*, PhD thesis, KTH, Stockholm, Sweden, 2018.

[4] Ternström S. Normalized time-domain parameters for electroglottographic waveforms. *J Acoust Soc Am*, 146, EL65-EL70, 2019.

*patelrir@indiana.edu

Effect of vocal intensity and fundamental frequency on cepstral peak prominence in women with and without voice disorders

Meike Brockmann-Bauser[1,2]*, Jarrad H. Van Stan[3], Marilia Carvalho Sampaio[1,2,4], Jörg E. Bohlender [1,2], Robert E. Hillman[3], Daryush D. Mehta[3]

[1] Department of Phoniatrics and Speech Pathology, University Hospital Zurich, Zurich, Switzerland
[2] University Zurich, Zurich, Switzerland
[3] Center for Laryngeal Surgery and Voice Rehabilitation, Massachusetts General Hospital, Boston, MA, USA
[4] Dept. Department of Speech, Language and Hearing Sciences, Federal University of Bahia, Salvador, Brazil

Keywords: smoothed cepstral peak prominence; voice diagnostics; voice intensity; fundamental frequency

Introduction

Cepstrum-based voice measures, such as smoothed cepstral peak prominence (CPPS), have been recently recommend for voice assessment to objectively describe overall dysphonia [1]. However, in vocally healthy adults, cepstral measures were influenced by voice sound pressure level (SPL) changes [2, 3]. Since it is unclear if similar effects hold in adults with voice disorders and how these interact with natural fundamental frequency (f_o) changes, this study examines voice SPL and f_o effects on CPPS in women with vocal hyperfunction and vocally healthy controls.

Methods

In a retrospective matched case-control study, 58 female voice patients 18–61 years of age (mean 27, SD 12.4) were paired with 58 vocally healthy women according to approximate age and occupation. The patient group comprised women diagnosed with phonotraumatic vocal hyperfunction associated with vocal fold nodules (n=39) or polyps (n=5), or non-phonotraumatic vocal hyperfunction associated with primary muscle tension dysphonia (n=14). All participants sustained the vowel /a/ at "soft", "comfortable", and "loud" conditions, and acoustic voice recordings were obtained at 10 cm distance. Voice SPL, f_o, and CPPS (dB) were computed from the acoustic voice recordings using Praat [4]. The effects of loudness condition, measured voice SPL, and f_o on CPPS were investigated with linear mixed models. Pairwise correlations among voice SPL, f_o, and CPPS were assessed using multiple regression analysis.

Results

Results (Figure 1) show that increasing voice SPL correlated significantly (p<.001) with higher CPPS in both patient ($r2=.54$) and normative control groups ($r2=.45$). f_o had statistically significant effects on CPPS (p<.001), with a weak relation for the patient ($r2=.02$) and control groups ($r2=.05$).

Discussion

While CPPS was moderately affected by the individual's voice SPL in women with and without voice disorders, the effects of f_o were considerably weaker. The direct relationship between CPPS and SPL in sustained vowels may be explained by a stronger harmonic source with increasing SPL. SPL related effects may be also present in connected speech samples of voice disordered adults, and also multiparametric indices incorporating CPPS [3]. Therefore, it is recommended to control for voice SPL variations in clinical assessments. This may be done by applying SPL-corrected values for CPPS or by taking care to elicit similar voice SPL levels throughout a patient's stage of treatment for robust CPPS comparisons. Other potentially confounding factors such as sex, age, and voice training status should be investigated in the future on patients across a larger spectrum of voice disorders.
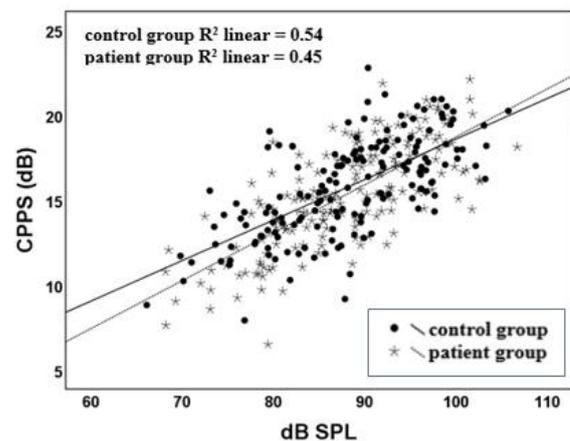
Figure 1: Relation between CPPS and voice SPL (dB SPL @10 cm). CPPS strongly increased with rising voice SPL in both the patient and control groups.

References

[1] Patel *et al.,* Am J Speech Lang Pathol*,* 27:887-905, 2018.
[2] Awan *et al.,* J Voice,26:670.e615-620, 2012.
[3] Phadke *et al.,* J Voice, Epub ahead of print, 17. Oct. 2018.
[4] Boersma *et al.,* PRAAT. 5.4.14, http://www.praat.org/, retrieved August 04, 2015.

corresponding author: meike.brockmann-bauser@usz.ch

Session #5 - Voice disorders and treatments

Airway reconstruction using posterior cricoid reduction for treatment of dysphonia

Elias Sundström*[1], Liran Oren[1], Alessandro de Alarcón[1,2]

[1]Department of Otolaryngology – HNS, University of Cincinnati, Cincinnati, Ohio, USA
[2]Division of Pediatric Otolaryngology– HNS, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, USA

Keywords: Airway reconstruction, dysphonia, FSI

Introduction

Nearly 50% of children who undergo airway reconstruction for subglottic stenosis have a significant risk of developing dysphonia due to postoperative posterior glottic diastasis [1]. Many of these patients develop alternative compensatory laryngeal structures as their primary sound source for voicing (e.g., the false vocal folds).

Two of the more common surgical interventions for correcting glottic diastasis are bilateral medialization thyroplasty and posterior cricoid reduction. Both procedures reduce the diastasis by medializing the glottic region. The Bilateral medialization thyroplasty technique is slightly different than the medialization technique that is typically used for unilateral vocal fold paralysis because anterior overcorrection must be avoided, and the posterior part of the implants cannot touch the arytenoid cartilage (and affect the vocal process). In posterior cricoid reduction, a vertical line (about 2 mm thick) is removed along the posterior cricoid, which medializes the glottic region when the lateral edges of the cricoid are reattached together.

The goal for the current study is to compare the pre- and postoperative airflow dynamics in a patient subjected to posterior cricoid reduction. Our hypothesis is that posterior cricoid reduction after airway reconstruction is an effective remedy for a compensatory speech disorder.

Methods

The geometry for the model was delineated from CT scans taken pre- and post-surgery in a patient who was diagnosed with dysphonia following airway reconstruction. The patient underwent posterior cricoid reduction and there is a three year time-lapse between pre- and post-scans. The scans covered the head and neck with a 0.5 mm slice thickness and obtained while the subject sustained the /e/ sound. The airflow was predicted using the incompressible Navier-Stokes equations. Simulation of tissue dynamics was done using fluid-structure interaction (FSI) [2]. The biomechanical properties of the tissue were based on literature data. The subglottal total pressure was a set to 6 cmH20 and atmospheric pressure was assumed at the mouth opening.

Results

Prior to surgery, the false vocal fold is subjected to a major pressure drop, which coincides with the location of the smallest cross-sectional area. As a result, the jet and associated acoustic source develop at the false vocal fold.

Post-surgery, undergoing cricoid reduction, the smallest cross-section was shifted to the true vocal fold. This modulates the major pressure drop to occur across the true vocal fold and the associated jet is shifted accordingly.
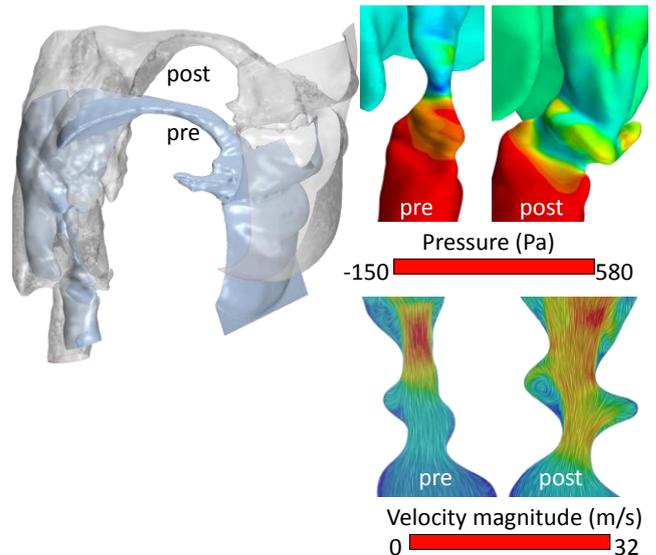
Figure 1: Isometric view of pre- and post-surgery airway reconstructions (left). Pressure and velocity distributions for pre and post inside the vocal tract (right). The flow goes from the subglottal inlet to the mouth opening.

Discussion

Prior to surgery, the patient was diagnosed with compensatory speech disorder using the false vocal folds, during phonation of the /e/ sound. With the posterior cricoid reduction, the smallest cross-sectional area was shifted to the true vocal folds. This relocates the acoustic source for phonation to the true vocal folds and correlates with a reduced compensatory speech disorder.

Acknowledgments

NA.

References

[1] de Alarcón, A., Zacharias, S., Oren, L., de Luzan, C. F., Tabangin, M. E., Cohen, A. P., Roetting, N. J., and Fleck, R. J. Endoscopic posterior cricoid reduction: A surgical method to improve posterior glottic diastasis. *The Laryngoscope.* 2019;129:S1-S9.

[2] Sundström, E., Jonnagiri, R., Gutmark-Little, I. et al. *Cardiovasc Eng Tech* (2019). https://doi.org/10.1007/s13239-019-00441-2

corresponding author e-mail: sundstes@uc.edu

32

*corresponding author e-mail: sundstes@uc.edu

Improved subglottal pressure estimation from neck-surface vibration in patients with voice disorders

Jon Z. Lin[1], Victor M. Espinoza[2,3], Katherine L. Marks[1,4], Matías Zañartu[2], Daryush D. Mehta[1,4,5]

[1]Center for Laryngeal Surgery and Voice Rehabilitation, Massachusetts General Hospital, Boston, MA, U.S.A.
[2]Department of Electronic Engineering, Universidad Técnica Federico Santa Maria, Valparaíso, Chile
[3]Department of Sound, Universidad de Chile, Santiago, Chile
[4]MGH Institute of Health Professions, Massachusetts General Hospital, Boston, MA, U.S.A.
[5]Harvard Medical School, Boston, MA, U.S.A.

Keywords: Subglottal Air Pressure; Ambulatory Voice Monitoring; Accelerometer; Inverse Filtering

Introduction

Prior work has shown promise for the noninvasive estimation of subglottal air pressure (Ps) using a linear model of the magnitude of neck-surface vibration during modal and non-modal voice production in vocally healthy speakers [1]. Subsequent work further developed this methodology by incorporating additional measures of vocal function from the neck-surface accelerometer (ACC) signal to achieve improved prediction of Ps during phonation [2]. This study expands on those studies by integrating these additional cepstral and glottal airflow measures to improve the prediction of Ps in patients with voice disorders, with the goal of tracking Ps in naturalistic, ambulatory settings.

Methods

Data were obtained from participants with voice disorders representing a variety of glottal conditions, including phonotraumatic vocal hyperfunction (PVH; associated with nodules/polyps), non-phonotraumatic vocal hyperfunction (NPVH; diagnosed as muscle tension dysphonia), and unilateral vocal fold paralysis (UVFP). Each patient repeated /p/-vowel syllables from loud-to-soft levels in multiple vowel contexts (/pa/, /pi/, and /pu/) and pitch conditions (comfortable, lower than comfortable, higher than comfortable) in their typical voice. Ps estimates were obtained via intraoral pressure (IOP) recordings during occlusive plosives using an intraoral catheter. Simultaneously, oral airflow was captured using a circumferentially vented pneumotachograph mask. Ps for each vowel was estimated by taking the average of IOP peaks preceding and following the vowel.

Two subject-specific, linear regression models were constructed to predict Ps using the ACC signal: Model 1 used the traditional ACC root-mean-square (RMS) magnitude, and Model 2 included additional ACC-based measures of vocal function. The additional measures in Model 2 included fundamental frequency (f_o), cepstral peak prominence (CPP), and glottal airflow parameters from subglottal impedance-based inverse filtering (IBIF) of the ACC signal [3]. Five-fold cross-validation within each patient's data set assessed the robustness of Model 1 and Model 2 performance using the root-mean-square error (RMSE) metric for each of the two regression models.

Results and Discussion

Each fold of the five-fold cross-validation exhibited a baseline prediction performance when Model 1 (ACC RMS alone) was used to predict Ps within each patient group. Improvements to Ps prediction performance (decreases in RMSE) were found when Model 2 added CPP, f_o, and the following glottal airflow measures: open quotient, speed quotient, normalized amplitude quotient, maximum flow declination rate, harmonic richness factor, peak-to-peak amplitude, and the difference between first and second harmonic amplitudes. Of note, similar model performance was achieved when the same glottal airflow-based IBIF measures were derived from the ACC signal using traditional airflow-based inverse filtering, thus showing promise for ACC-only prediction of Ps in clinical populations.

Patient group	Model 1 RMSE	Model 2 RMSE	ΔRMSE (cm H_2O)	ΔRMSE (%)
PVH	2.31 (1.06)	1.78 (0.65)	−0.53 (0.50)	−20.92 (10.87)
NPVH	2.40 (1.02)	2.07 (1.09)	−0.33 (0.24)	−15.81 (12.38)
UVFP	2.36 (0.87)	2.07 (0.81)	−0.29 (0.31)	−11.82 (13.59)

Table 1: Improvements in Ps prediction performance in terms of mean (standard deviation) root-mean-square error (RMSE) within each patient group, comparing the accelerometer RMS-only linear regression model (Model 1) with a multiple linear regression model (Model 2) that incorporated accelerometer-based measures of CPP, f_o, and glottal airflow measures derived using subglottal impedance-based inverse filtering. Change in (Δ) RMSE also reported in cm H_2O and as a percentage with Model 1 as reference.

Acknowledgements

References

[1] Marks *et al*, J Speech Lang Hearing Res, 62:3339–3358, 2019.
[2] Lin *et al*, IEEE J Sel Topics Signal Process, In Press, 2019.
[3] Zañartu *et al*, IEEE Trans Audio Speech Lang Process, 21:1929–1939, 2013.

Three-dimensional vocal fold structural change due to implant insertion in medialization laryngoplasty

Dinesh Chhetri, Liang Wu, Zhaoyan Zhang

Department of Head and Neck Surgery, UCLA School of Medicine, Los Angeles, CA, USA

Keywords: voice inversion; machine learning; voice simulation

Introduction:

Previous investigations suggest that soft thyroplasty implants lead to superior acoustics compared to stiff implants in the treatment of glottic insufficiency. The goal of this study was to quantify the three-dimensional structural change of the vocal folds due to implant insertion in medialization laryngoplasty, and its potential effect on voice production.

Methods:

Medialization laryngoplasty were performed in excised human larynges using soft (silicone) and stiff (Silastic) implants. MRI images of the larynges were obtained and used to reconstruct the three-dimensional laryngeal geometry with and without implant insertion.

Results:

Implant insertion deformed the vocal folds into a thin layer wrapped around the implant. The medial-lateral dimension of the vocal folds was significantly reduced from about 4 mm to 1 mm, and the vocal folds were stretched in the coronal plane by about 70%. Insertion of stiff implants also led to noticeable soft tissue tears and gaps between the implant and the vocal folds. In contrast, soft implants were better able to mold into the vocal folds with less tears or gaps.

Conclusions:

With implant insertion, the implant assumed the role of the body layer in the coupled vocal fold-implant system, whereas the original TA muscle and cover layer became the new cover layer. Use of implants with stiffness comparable to that of the vocal folds is recommended because the degree of medialization can be adjusted without much negative effects on phonation frequency, phonation threshold pressure, or vibration amplitude.

corresponding author e-mail: dchhetri@mednet.ucla.edu

Duration of biodynamic changes associated with water resistance therapy

Matthias Echternach[1], Marie Köberlein[1], Marco Guzman[2], Anne Maria Laukkanen[3], Bernhard Richter[4], Marie-Anne Kainz[1], Michael Döllinger[5]

[1]Division of Phoniatrics and Pediatric Audiology, Department of Otorhinolaryngology, Munich University Hospital (LMU)
[2]Department of Communication Sciences and Disorders, Universidad de los Andes, Chile. Santiago, Chile
[3]Speech and Voice Research Laboratory, Faculty of Social Sciences, Tampere University, Tampere, Finland
[4]Institute of Musicians' Medicine, Freiburg University Medical Center and Medical Faculty, Freiburg University
[5]Division of Phoniatrics and Pediatric Audiology at the Department of Otorhinolaryngology University Hospital Erlangen

Keywords: semi-occluded voice training exercise (SOVTE) – high-speed imaging – EGG

Introduction

In current voice research, there is a growing understanding of how semi-occluded vocal tract exercises contribute to an increase of vocal efficiency. However, it has not yet been clarified for how long the effects last.

Material and Methods

Eight vocally healthy subjects were asked to sustain a phonation on the vowel /i/, fundamental frequency of 250Hz (females) or 125Hz (males), and at a comfortable loudness. During phonation the subjects were simultaneously recorded with transnasal high speed videoendoscopy (HSV, 20.000fps), electroglottography, and audio signals. After that, these subjects performed a WRT for 10 minutes (tube of 30cm length, 5cm below the water surface). Repeated measurements of sustained phonation were performed 0, 5, 10, 20 and 30 minutes after exercising. From the HSV material the Glottal Area Waveform (GAW) was segmented and GAW parameters were computed.

Results

There were strong inter-individual differences concerning the courses of the different measures after WRT. In general, directly after WRT there was a lowering of the GAW derived Period Perturbation Quotient, a lowering of the closing quotient and an increase of the sound pressure level in comparison to the pre intervention measurement. However, only 5 minutes post WRT there was no longer a clear difference compared to the pre intervention. Other values such as open quotients exhibited no evident change directly after the intervention.

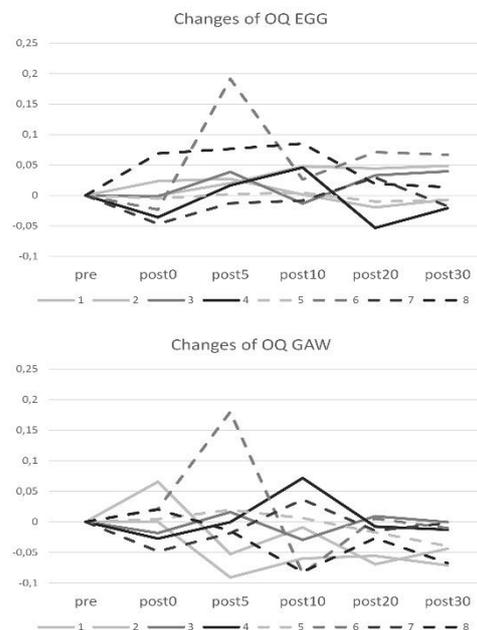*Figure 1: Changes of the Glottal Area Waveform (GAW) and the electroglottographical (EGG) Open Quotient (OQ) for all subjects*

Conclusions

WRT showed strong inter-individual effects concerning the courses of the evaluated measures. General tendencies of some measures directly after the intervention showed a brief effect of only a few minutes

..

Session #6 - Fluid-structure-sound interaction in the vocal tract

Detecting sound modes in a vocal tract model using Particle Image Velocimetry and Proper Orthogonal Decomposition

Charles Farbos de Luzan[1]*, Liran Oren[1], Alexandra Maddox[2], Ephraim Gutmark[2], Sid Khosla[1]

[1] Department of Otolaryngology – head and neck surgery, University of Cincinnati, Cincinnati, Ohio
[2] Department of Aerospace Engineering – CEAS, University of Cincinnati, Cincinnati, Ohio

Keywords (Maximum 4): PIV; POD; NAH

Introduction

Prior work used near-field acoustic holography [1] to identify sound modes produced by an electrolarynx in a mechanical model of a vocal tract. The same method was applied to a live human in order to map the sound modes during voicing of different sounds. Increasing the near- and far- fields occlusions in the vocal tract affected the sound mode near the source. It showed that in the case of "nasal" sounds the main mode was more shifted toward the nostrils of the subject. The current experiment used particle image velocimetry (PIV) to measure the velocity fields in a mechanical vocal tract placed above an excised canine larynx. The goal is to show the correlation between the results obtained from NAH and POD applied to the velocity fields, in order to predict the sound modes in a vocal tract model using PIV.

Methods

One canine larynx was excised after the animal was euthanized. PIV measurements were obtained in three regions above the glottis, namely in the supraglottal region up to the false vocal folds occlusion (FVF), between the FVF and the far-field occlusion, representing the mouth occlusion (MC), and above the MC. Measurements were obtained at low and high subglottal pressures, relatively to the phonation threshold pressure. The occlusion at the FVF level was varied (no FVF, 7mm gap, and 3mm gap), and the occlusion at the MC was varied from existing to none (open). Proper orthogonal decomposition was applied to the measured fields and the energy of each mode was compared to the other configurations, in order to evaluate the mode location.

Results

Figure 1 shows the energy for the modes 1 and 2 at the three locations for the 6 considered configurations. Energy levels decrease at the FVF level when reducing the FVF gap. Adding the occlusion at the mouth level increases the first mode near the source. These effects are more visible at low subglottal pressure (a few cmH2O above PTP). At high subglottal pressure, an occlusion at the mouth increase the energy level at the mouth in the presence of FVF occlusion.
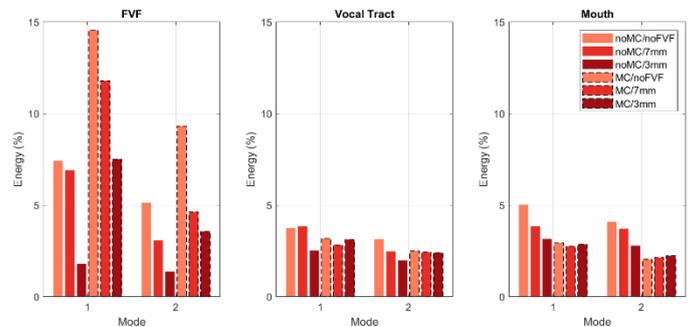
Figure 1: Energy for modes 1 and 2 at low subglottal pressure.

Figure 2 compares the strength of mode 1 obtained from NAH (left side of vignettes) and POD (right side). With an open vocal tract, POD shows a strong mode 1 through it and above, while adding MC pushes most of the energy closer to the FVF.
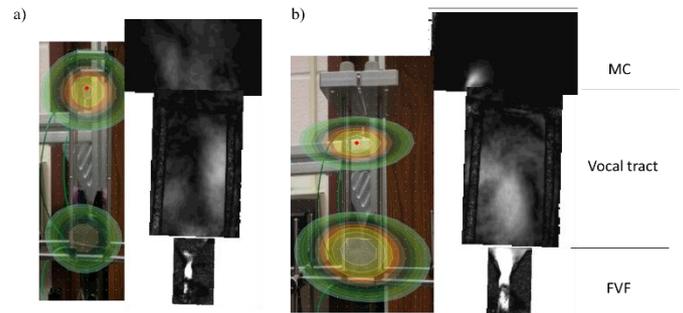
*Figure 2: Contour plots showing mode 1 for **a.** 3mm gap FVF and no MC, and **b.** 3mm FVF gap and MC.*

Discussion

Based on the velocity fields, varying the occlusion at the FVF and mouth affects the energy levels near the source. These preliminary results seem to confirm what was previously observed using NAH. Using PIV yields spatial dynamics of the supraglottal flow fields, and more in-depth comparison of NAH could make possible the use of NAH as a diagnostic tool for voice disorder.

References

[1]. Maynard, Julian D., Earl G. Williams, and Y. Lee. "Nearfield acoustic holography: I. Theory of generalized holography and the development of NAH." *The Journal of the Acoustical Society of America* 78, no. 4 (1985): 1395-1413.

Aeroacoustic simulation on a simplified vocal tract model
with tongue movement for the articulation of [s]

Tsukasa Yoshinaga[1]*, Kazunori Nozaki[2], Hiroshi Yokoyama[1], Akiyoshi Iida[1]

[1]Department of Mechanical Engineering, Toyohashi University of Technology, Toyohashi, Aichi, Japan
[2]Division of Medical Informatics, Osaka University Dental Hospital, Suita, Osaka, Japan

Keywords: Aeroacoustics; Large Eddy Simulation; Vocal Tract Model; Sibilant Fricatives

Introduction

Sibilant [s] is one of fricative consonants, and known to be pronounced by forming a jet flow and its aeroacoustic sound in the front part of vocal tract [1]. When the [s] is produced, a constricted flow channel is formed by elevating a tongue tip towards an alveolar ridge.

The production mechanisms of [s] have been investigated by several researchers. Hamlet *et al.,* [2] used electro-magnetic sensors and showed that the tongue contact on the hard palate occurs after the appearance of fricative noise, whereas the duration of the sound is longer than the duration of the tongue contact. In our group, to clarify the relationship between the flow and sound generation, a simplified vocal tract model of [s] was constructed based on medical images, and large eddy simulation (LES) was conducted [3]. In addition, effects of tongue movement speed on the articulation of /s/ were investigated by experiments using the movable tongue model [4]. However, the mechanisms of jet flow and source generation during the tongue movement are still unclear. Therefore, in this study, the LES is conducted on the simplified vocal tract model with movable tongue to clarify the relationship between the flow and sound generation in the articulation of [s].

Methods

The simplified vocal tract model is depicted in Fig. 1. The tongue model ascended and descended by 3 mm with tongue speed 40 mm/s from the position of [s] with constant flow rate 18 L/min. The three-dimensional Navier-Stokes equations were solved by finite difference method and volume penalization method which allows the wall boundary to move in the vocal tract.
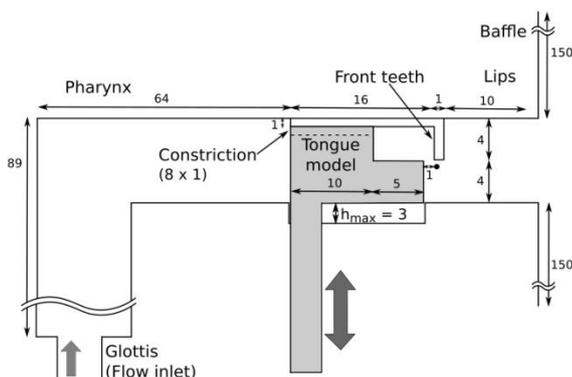
Figure 1: Simplified vocal tract model for the articulation of /s/.

Results and Discussion

The velocity distribution on mid-sagittal plane of the simplified vocal tract model and iso-surfaces of second invariant of velocity gradient tensor during tongue ascent are shown in Fig. 2. Results showed that the turbulence started occurring and the jet flow formed the vortex tubes which becomes the sound source during the tongue ascent. This suggests that the proposed simulation is applicable for the investigation of fricative articulations.
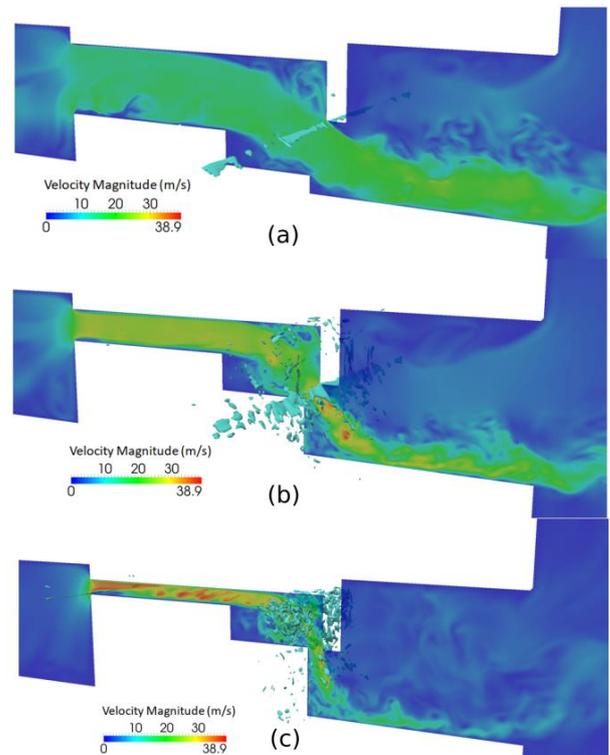
Figure 2: Velocity distribution on mid-sagittal plane of the simplified vocal tract model and iso-surface of Q during tongue elevation (a)h = 0 mm, (b)h = 2.16 mm, (c)h = 3 mm.

Acknowledgements

References

[1] Stevens, K. N., *Acoustic Phonetics* (The MIT Press), 1998.
[2] Hamlet *et al.*, J. Acous. Soc. Am., 65:1276-1285, 1979.
[3] Yoshinaga *et al.*, Phys. Fluid. 30:035104, 2018.
[4] Yoshinaga *et al.*, PLoS One 14:e0223382, 2019.

corresponding author e-mail: yoshinaga@me.tut.ac.jp

Phonation energy budget from high-fidelity aeroelastic-aeroacoustic simulations

Lucy Zhang[1]*, Feimi Yu[1], Michael Krane[2]

[1]Mechanical, Aerospace & Nuclear Engineering, Rensselaer Polytechnic Institute, Troy, NY, USA
[2]Applied Research Laboratory, Penn State University, State College, PA, USA

Keywords: aeroeelastic-aeroacoustic simulation, phonation, energy budget, IFEM

Abstract

A rigorous accounting of phonation energy utilization is presented using high-fidelity computer simulations. Simulation results are used to compute terms of the integral energy equation for the volume containing air in the larynx. Flow work terms are decomposed to clarify power transfer mechanisms. Laryngeal acoustic efficiency is presented.

Introduction

Using a control volume analysis, the energy budget of air motion in the larynx has been defined [3]. Such an accounting reveals the power transfer mechanisms from the pulmonary system, vocal fold vibration, glottal jet dissipation, and energy exchanges between the airflow in the larynx and the acoustic fields in the trachea and the vocal tract.

Methods

To evaluate the energy budget, an aeroacoustic-aeroelastic simulation is setup using a swept-ellipse multilayer model [1,3]. The simulation is performed with the immersed finite element method [2]. A pair of vocal folds is placed within the vocal tract, represented by a stright uniform cross-section duct. The mouth opening is represented by an extended large region with non-reflective boundaries, shown in Fig. 1.
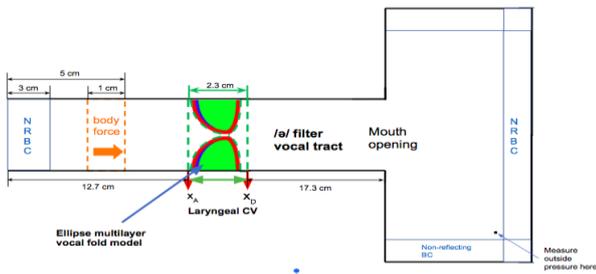
Figure 1: High-fidelity simulation setup of a pair of elliptical vocal folds placed in a vocal tract.

Results and Discussion

The power flows for the terms over several vibration cycle is shown in Fig. 2. The net pressure work, $2(p_A^+ Q_A - p_D Q_D)$, is balanced with the radiated wave pressure $\frac{\rho c}{S}(Q_D^2 + Q_A^2)$, and other outputs and losses $(\dot{W}_{VF}, \dot{W}_v, \dot{KE}, \dot{PE}, \dot{W}_f)$.
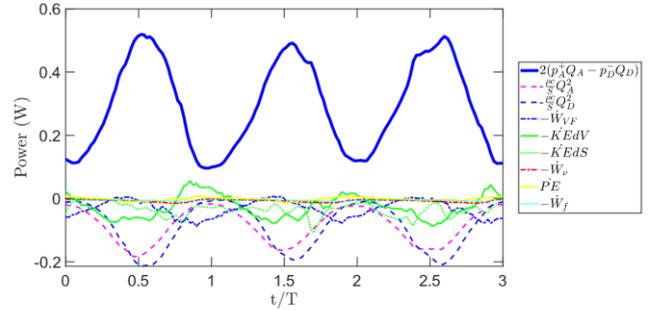
Figure 2: Power flow from the net pressure work (solid blue) to radiated pressure (dotted blue and dotted cyan) and other outputs and losses with input pressure of 1kPa.

Three pressure inputs (0.8, 1.0, 1.2KPa), forced/unforced flow symmetry, and infinite duct without mouth, are simulated for comparison of the energy utilization, shown in Fig.2. Overall, the acoustic output is within 20% of the pressure work input, the energy loss to viscous dissipation and skin friction, and potential energy to compress air are insignificant. Asymmetric cases generate higher work input, output, and losses. As the pressure input increases, the generated work is also higher.
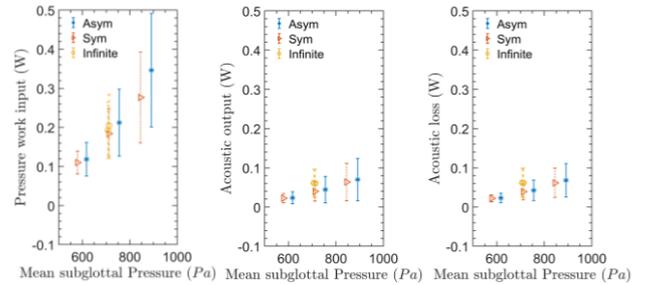
Figure 3: Pressure power input, acoustic output and losses with standard deviations vs. three pressure inputs, forced/unforced flow symmetry, and infinite duct without mouth opening.

Acknowledgements

References

[1] McPhail et al. Aeroacoustic source characterization in a physical model of phonation, JASA 2019.
[2] Yang et al. Fully-coupled aeroelastic simulation with fluid compressibility, CMAME 2017.
[3] Krane et al. Voice energy utilization and efficiency, ICVPB 2018.

corresponding author e-mail: zhanglucy@rpi.edu

corresponding author e-mail: zhanglucy@rpi.edu

Toward aeroelastic-aeroacoustic phonation model validation

Jeff Harris[1]*, Adam Nickels[1], Michael Krane[1], Paul Trzcinski[1,2], Zachary Yoas[1], Faith Beck[1], Benjamin Beck[1]

[1]Applied Research Laboratory, Pennsylvania State University, State College, PA, USA
[2]Aerospace Engineering Department, Pennsylvania State University, State College, PA, USA

Keywords: Aeroacoustics; digital image correlation

Introduction

Standards for validation of models and simulations have been implemented in the defense and energy industries, and for FDA certification of medical devices [1-4], but have not yet been completely implemented in phonation. Standard practice for validation dictates that system response quantities (such as velocity, pressure, vibration, etc.) over a wide range of a difficulty spectrum would provide the most useful dataset for validation purposes [5]. The data also require a quantification of uncertainty (both experiment and simulation) to aid in the validation process. Lastly, the measurements should be non-intrusive, using optical techniques where possible.

To provide a validation dataset that meets the stringent requirements described in [5], system response quantities (SRQs) and boundary conditions (BCs) are measured simultaneously. Simultaneous measurements of acoustic pressure, transglottal pressure, volume flow, vocal fold surface motion are reported using the Penn State Upper Airway Model (PSUAM) [6, 7]. This simplified model provides a fully-quantifiable and controllable test apparatus. The as-built dimensions, material properties, and eventually the inflow conditions are or will be measured as the system BCs. In particular, the frequency-dependent modulus of loss factor of the vocal fold are characterized using dynamic mechanical analysis of the rubber layers [8]. These are used as direct inputs to simulations with the intent of validation. Using those BCs, the simulations compute the same SRQs as those that are measured in the experiment for comparison.

Experiment

The PSUAM [6-7] with a multilayer swept-ellipse vocal fold (VF) model, in a hemilarynx configuration was modified to allow for the simultaneous acquisition of pressures, glottis motion, and the velocity field. The setup is shown in Figure 1 showing the PSUAM, camera layout, and the location of the pressure sensors. The modification was merely to use half of the glottis rather than a symmetric VF with two sides.

Pressure and volume flow measurements were conducted as described in [6-7], and VF surface motion was acquired by imaging the surface with three high-speed cameras. The surface motion was computed from the video.

The flow rate of the airway was set to several different magnitudes, with data acquired at 10kHz, sampled for 10 seconds.

Results

Results will include the VF surface motion, pressure and acoustic response, and flow rate with uncertainties for the purpose of validation of aeroelastic-aeroacoustic models.
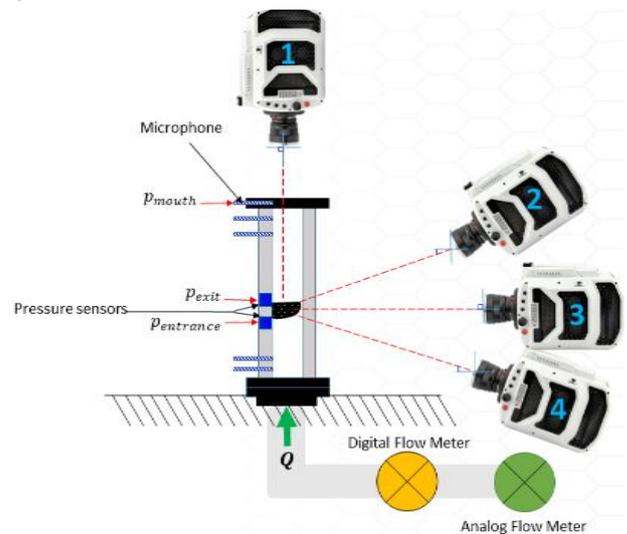
Figure 1: The PSUAM model with a hemilarynx. Camera 1 gives direct video of glottal gap, cameras 2-4 for DIC of VF surface. The flow meters, microphones, and pressure sensors are shown.

Acknowledgements

References

[1] Lee, J., Rygg, A., et al. "Fluid-Structure Interaction Models of Bioprosthetic Heart Valves: Initial In Vitro Experimental Validation." (2019). (EngrXivi)

[2] Craven, B. A., et al. "Steady Flow in a Patient-Averaged Inferior Vena Cava—Part II: Computational Fluid Dynamics Verification and Validation." Cardiovascular eng. and tech., 9(4), 2018

[3] Hariharan, P., et al. "Inter-laboratory characterization of the velocity field in the FDA blood pump model using particle image velocimetry (PIV)." Cardiovascular eng. and tech., 9(4), 2018.

[4] Malinauskas, R., et al. "FDA benchmark medical device flow models for CFD validation." Asaio Journal 63(2) (2017).

[5] Oberkampf, W. L., and Roy, C. J. *Verification and validation in scientific computing*. Cambridge University Press, 2010.

[6] Campo, L., "The effect of vocal fold geometry on the fluid structure acoustic interactions in an experimental model of the human airway," M.S. Thesis, Penn State Univ, 2012.

[7] McPhail, et al. "Aeroacoustic source characterization in a physical model of phonation." *JASA* 146(2) 1230-1238, 2019.

[8] ASTM Std. D4056-12, 2012, "Standard Practice for Plastics: Dynamic Mechanical Properties," ASME International, 2012.

Pharyngeal air pressure along the vocal tract during vowels and semi-occluded vocal tract exercises: A pilot study using high-resolution pharyngeal manometry

Jesse Hoffmeister[1,2]*, Christopher Ulmschneider[1], Corinne Jones,[3] Michelle Ciucci[1,2,4], Timothy McCulloch[2]

[1]Department of Communication Sciences and Disorders, University of Wisconsin-Madison, Madison, Wisconsin, USA
[2]Department of Surgery-Division of Otolaryngology, University of Wisconsin-Madison, Madison, Wisconsin, USA
[3]Department of Neurology, University of Texas-Austin, Austin, Texas, USA
[4]Neuroscience Training Program, University of Wisconsin-Madison, Madison, Wisconsin, USA

Keywords: manometry; pharyngeal; pressure; phonation

Introduction

The study of air pressure in the vocal tract is central to the study of voice. Vocal tract air pressure is associated with phonatory efficiency, and changes to air pressure impact the vocal acoustic signal[1-3]. As such, accurate air pressure measurement is important for research and clinical practice[4]. Current methods of air pressure measurement in the vocal tract are generally limited to the anterior oral cavity and to a restricted number of phonatory gestures [4,5]. This is problematic because air pressure at different locations in the vocal tract may vary. This feasibility study proposes a novel method of measuring air pressure simultaneously at multiple levels of the lower vocal tract using high-resolution pharyngeal manometry (HRM).

Methods

Two males with prior singing training underwent HRM. A catheter was passed transnasally and air pressure was measured simultaneously at 5 points 1cm apart in the lower vocal tract between the velopharyngeal port and the upper esophageal sphincter during multiple phonatory gestures, including semi-occluded vocal tract exercises[6]. Air pressure was compared among locations in the vocal tract and among phonatory gestures. Pressure measurements obtained in the current study were compared to those reported in similar studies that measured air pressure in the anterior vocal tract. Descriptive statistics were calculated for pressure data at each of the 5 sensors for each task, as well as for each task combined across sensors. Means, standard deviations and confidence intervals from the current study were then compared to those reported in the extant literature [7,8].
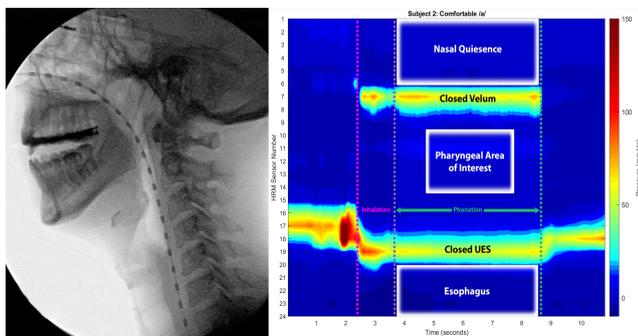
Figure 1: Left: Sample Fluoroscopic image of transnasal catheter placement during HRM. Right: spatiotemporal plot of pressure differentials during HRM. Warmer colors represent increasing pressure. Pharyngeal Area of Interest demonstrates the location of the sensors on the y-axis from which data were analyzed.

Results

HRM was well-tolerated in both subjects. Pressures were non-uniform across lower vocal tract locations, with increased pressure observed immediately distal to the velopharyngeal port. Air pressure averaged across all sensors did not consistently fit into confidence intervals for air pressures taken from the anterior oral cavity during the same tasks in the extant literature. Pressure relationships among sensors differed by type of phonatory task performed. Pharyngeal air pressure increased with increasing vocal tract semi-occlusion.
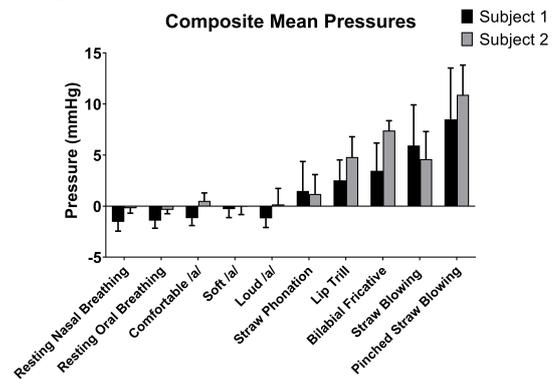
Figure 2: Mean pressure averaged across all five sensors for each subject. Error bars indicate standard deviation.

Discussion

HRM is capable of measuring pressure simultaneously at multiple levels of the lower vocal tract during phonatory tasks with high spatial and temporal resolution, providing rich data with which to analyze the physical properties of voice production. This study demonstrates inconsistencies in pressure at different locations in the lower vocal tract during semi-occluded vocal tract exercises. This instrumental assessment may be useful in future assays exploring differences in air pressure in the lower vocal tract between normal and disordered populations.

References

[1] Titze, J Acoust Soc Am. 123:2733-2749, 2008
[2] Bartholomew J Acoust Soc Am. 6:25-33, 1934
[3] van Houtte et al, J Voice. 25(2):202-207, 2011
[4] Patel *et al, Am J Speech Lang Pathol*. 27:887-905, 2018
[5] Smitheran *et al*, J Speech Hear Disord. 46:138-146, 1981
[6] Titze, J Speech, Lang Hear Res. 49:448-459., 2006
[7] Maxfield et al, Logop Phoniatr Vocology. 40(2):86-92.2015
[8] Guzmán *et al*, J Voice. 30(6):759.e1-759.e10, 2016

*hoffmeister@surgery.wisc.edu

Poster Multi-sessions

Mathematical Modeling of the Medial Surface of the Vocal Fold for the Study of Chest and Falsetto Registers

Douglas Blake[1*], Ingo Titze[2], Eileen Finnegan[1]

[1]Communication Sciences and Disorders, University of Iowa, Iowa City, Iowa, USA
[2]National Center for Voice and Speech (NCVS), University of Utah, Salt Lake City, Utah, USA

Keywords: Modeling; Registers; Biomechanics; Singing

Introduction

Experimental data suggest that the change of laryngeal register between chest and falsetto is set by the level of activation of the thyroarytenoid muscles, with increased thyroarytenoid activation causing medial surface bulging which characterizes chest register [1]. This project explored whether medial surface bulging resulting from the activation of the thyroarytenoid muscle could cause change in the spectral slope of the acoustic output.

Methods

The geometry for the finite element model was based on the typical male geometry defined in VoxInSilico [2], a voice simulation program written by Drs Ingo Titze and Fariborz Alipour. Modeling was performed in FEBio using a transverse-isotropic Mooney-Rivlin model for the muscle tissue and Neo-Hookean models for the mucosa and ligament. The length of the vocal fold was adjusted by motion of the thyroid cartilage relative to a fixed posterior surface. The muscle tissue was divided into vocalis and muscularis portions to allow variation in fiber orientation and activation.
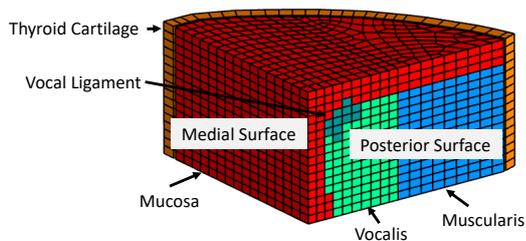
Figure 1: FE Model showing muscle divisions and vocal ligament

The FE model was solved for 50 cases of muscle activations that covered the expected range of muscle activations for the chest and falsetto registers. The displacements of the medial surface were extracted from each of those solutions. The extracted medial surface displacement data was regressed into a polynomial function of muscle activations and position on the medial surface. This 2-D polynomial was inserted into VoxInSilico to calculate the shape of the medial surface. A sample chest and falsetto case were then run in VoxInSilico and the spectral slopes of the acoustic output calculated.

Results

The expectation was that the FE model would show more bulging for the chest than for the falsetto case, and that the chest case would have a shallower spectral slope. The deformed shape of the vocal fold and the cross-section at the posterior edge are shown for the falsetto case in Figure 2 and for the chest case in Figure 3. The spectral slopes for the chest and falsetto case are shown in Table 1.
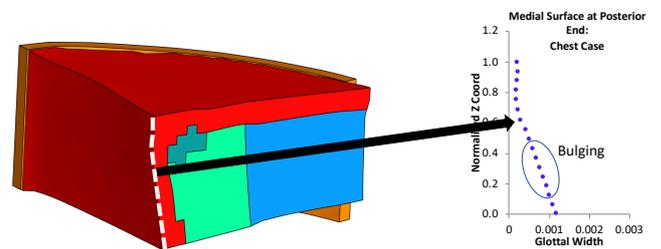
Figure 2: FE Model for chest (left) and posterior cross section plot (right) shows bulging.
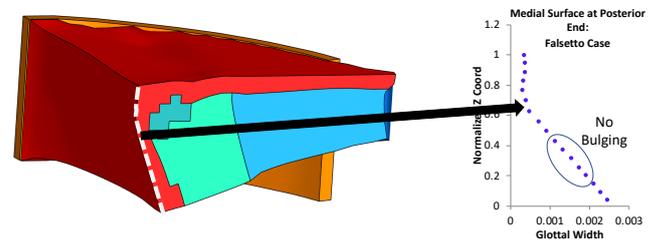
Figure 3: FE Model for falsetto (left) and posterior cross section plot (right) shows no bulging.

	Chest	Falsetto
Spectral Slope	-11.9 dB/oct	-19.2 dB/oct

Table 1: Spectral Slopes for the Chest and Falsetto Cases.

Discussion

The higher muscle activation of the chest case did result in more medial surface bulging and a shallower spectral slope. Future work will include identifying additional chest and falsetto cases to further exercise the regression model.

References

[1] Hirano, M. *et al*, Folia Phoniatr, 22:1-10, 1970.
[2] VoxInSilico (Version 7.3) [Computer Software]. Salt Lake City, UT: National Center for Voice and Speech

*douglas-blake@uiowa.edu

Singing Register Differences in Vocal Fold Oscillations Observed Across Three Octaves Through Laryngeal High-Speed Videoendoscopy

Hugo Lehoux[1]*, Lisa Popeil[2]†, Jan G. Švec[1]‡

[1]Voice Research Lab, Dept. Biophysics, Faculty of Science, Palacký University Olomouc, Czech Republic
[2]Voiceworks, Los Angeles, California, USA

Keywords: vocal registers; vocal fold oscillation; vocal range; high-speed videoendoscopy

Introduction

Because the definition of voice registers is mostly based on perceptual characteristics and singers' interior sensations, understanding their physiological differences requires analyzing quantifiable parameters. Therefore, the objective of this study is to look for relevant vocal-fold (VF) vibration parameters which are able to discriminate between the "modal" register (M, speech-like voice sound) and the "head" register (H, flutey sound) [1,2]. This study investigated a professional singer who demonstrated a unique ability to choose between the two registers regardless of pitch throughout her entire singing range [3].

Methods

The professional singer (LP) performed short phonation on all pitches from C3 = 131 Hz to C#6 = 1109 Hz, first in M then in H, using the same 'laryngoscopic vowel' reminding of [e:]. The VF vibration was registered using a high-speed camera connected to a rigid 90° endoscope, at 7200 frames per second (fps). Some of the higher pitches were also registered at 13600 fps. A condenser omnidirectional microphone registered the radiated sound and an ElectroGlottoGraph (EGG) registered the VF contact area. PhonoVibroGrams (PVG) and digital kymograms were computed and used for analysis.

Results and Discussion

Visual observations of the high-speed videos, kymograms, glottal area waveforms (GAW), PVG, and EGG signals suggest that the whole range can be separated in a lower (C3 = 131 Hz to G4 = 392 Hz) and a higher (G#4 = 415 Hz to C#6 = 1109 Hz) subranges. In the lower subrange, clear differences in the VF contact duration were observed in EGG and kymographic signals revealing on VF adduction changes between M and H. Further analysis of the Normalized Amplitude Quotient (NAQ) from GAW (suggesting changes in VF lower margin vibration amplitudes), the closed quotient from PVG, and the sharpness of the lateral peaks from the kymograms (reflecting VF vertical phase differences) confirmed these differences in VF vibratory patterns. In the higher range, clear differences between M and H were found only for the NAQ parameter.
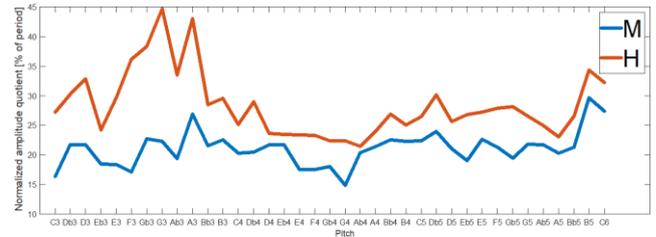
Figure 1: Normalized Amplitude Quotient (NAQ) derived from GAW showing consistent register differences across all pitches. Smaller NAQ suggests larger vibratory amplitudes of the lower VF margin in M register.
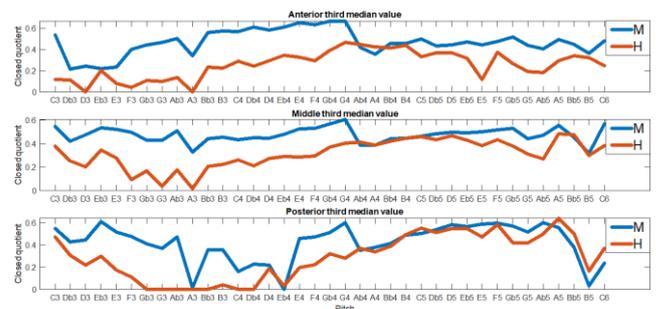
Figure 2: Closed Quotient derived from PVG Contour revealing on M/H register differences in membranous (anterior, middle) and cartilaginous (posterior) VF adduction.

In conclusion, the analysis of vocal fold vibratory patterns reveals there are parameters allowing differentiating both registers. The M/H differences present themselves more clearly in the lower pitch range of voice, however.

Acknowledgments

References

[1] Roubeau B, Henrich N, Castellengo M (2009) Laryngeal Vibratory Mechanisms: The Notion of Vocal Register Revisited. J Voice 23: 425-438.

[2] Svec JG, Sundberg J, Hertegard S (2008) Three registers in an untrained female singer analyzed by videokymography, strobolaryngoscopy and sound spectrography. J Acoust Soc Am 123: 347-353.

[3] Echternach M, Popeil L, Traser L, Wienhausen S, Richter B (2014) Vocal tract shapes in different singing functions used in musical theater singing-a pilot study. J Voice 28: 653.e1-653.e7.

*hugo.lehoux01@upol.cz, †lisapopeil@mac.com, ‡jan.svec@upol.cz

The balloon in the box model; exponential factors in voice control

Peter Pabon[1]*

[1]Institute of Sonology, Royal Conservatoire, The Hague, The Netherlands

Keywords: VRP; spectrum balance; subglottal pressure; statistical mechanics;

Introduction

An essential element in the mapping of spectral voice metrics over the VRP [1], is the exclusive use of logarithmic scales. This means that when a linearized dependency is observed as a function of the (logarithmic) fundamental frequency or the SPL (the logarithm of the acoustic pressure), then it will actually represent a constant *exponential factor* in the relationship. For instance, the color maps that show the distribution of the spectrum balance (SB) metric over the interior of the VRP (Figure 2), display sections with constant gradients, e.g., constant exponential factors. It is possible to zoom into these maps and to study the linearization of the SB-SPL relationship in detail, at selected vertical bands of constant fundamental frequency. Remarkably, the exponents that are associated with the SB-SPL linearization vary largely between individuals, but are very reproducible within individuals. Moreover, seen over the total SPL range, the observed SB-SPL dependency, in some cases, exhibits several regimes, with different exponential factors depending on the SPL interval. The SPL where such regime changes might occur, varies again between individuals, but appears to be very reproducible within individuals [2]. If we assume that the voice production mechanism is largely comparable between individuals, how can there be such diversity in (1) the sloping, (2) the offset for the SB-SPL slope, and (3) in the appearance and alignment of different regimes? What is the origin of these regimes?

Thesis

The exponential relationships that are observed with this mapping paradigm would not be so systematic if they were not an expression of an underlying generating principle that follows the same paradigm. In this paper, the idea is explored that constant proportional dependencies (constant exponents) are the shared factors in the parametric control of the voice.
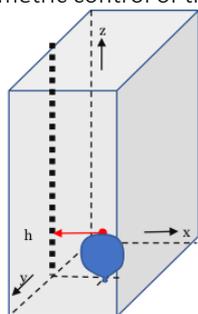
Figure 1, balloon in the box model.

To illustrate how a single physical system can exhibit regime changes, the model of a balloon in a box is used as an analogy (Figure 1). I will discuss the embedding of this model in statistical mechanics, the transformation that model parameters go through when going from the micro-scale to the macroscale, and the implications that such a model has for the dependency between subglottal pressure and acoustical pressure. VRP maps that show the different regimes of the SB metric for different voices will be reviewed to motivate the idea.
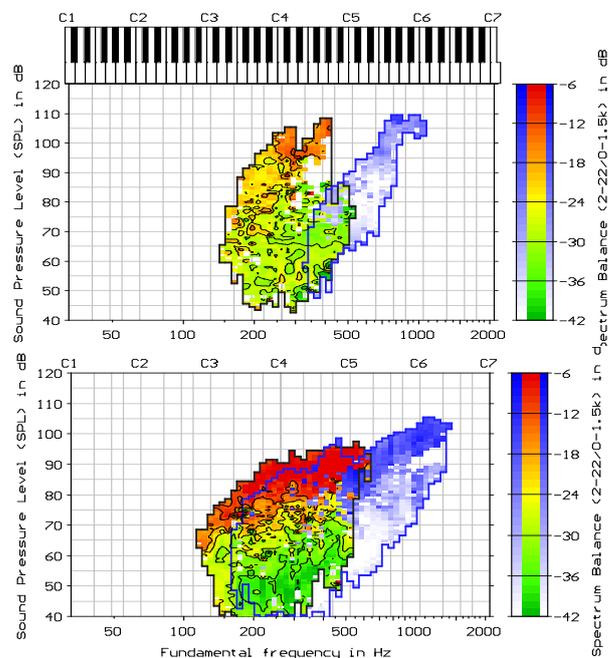
Figure 2: Maps of the Spectrum Balance metric over the VRP for the modal voice range (M1) for two different untrained female voices. Taken from [1] to demonstrate how different these distributions can be in offset, range and incline for different individuals.

References

[1] Pabon P, Ternström S, Feature maps of the acoustical spectrum of the voice, J Voice, 2018. https://doi.org/10.1016/j.jvoice.2018.08.014.

[2] Patel, R. Ternström S, Non-invasive evaluation of vibratory kinematics of phonation in children. Paper abstract submitted to ICVPB 2020.

Common base of western and non-western scales derived from vocal tract resonances

Patrick Hoyer [1]* and Louisa Traser[2]

[1]Fraunhofer Headquarters, Munich, Germany
[2]Institute of Musicians' Medicine, University Medical Center – University of Freiburg, Faculty of Medicine, Germany

Keywords: Vocal Ergonomics; Musical Ethnology; Formant Tuning; Intervals

Introduction

The human voice is considered predominant in the emergence of music as a whole and vocal capabilities in combination with hearing abilities are regarded to support the emergence of intervals in any cultural setting [1].

Physical properties of the vocal tract give rise to relations that have been linked to intervals on an auditory level [2]. The present work focusses on vocal ergonomics due to vocal tract resonances as possible driving force for the emergence of intervals or scales.

Methods

The spectra and sound pressure level of an externally stimulated 3D vocal tract model [3] during a pitch glide was recorded and analysed. The experimental data are compared via a semi-empirical simulation procedure first introduced by Fant [4].

Background, Results and Discussion

At comfortable singing range close to average speaking frequency the vocal tract resonances f_{Rn} are considered to be constant at a given vowel at first approximation. The average speaking frequency is linked to the first and second vocal tract resonances [4]. A vocal tract resonance supporting a harmonic will amplify most effectively at multiples of the fundamental frequency f_o. Since vocal tract resonances are much higher than any f_o used during speaking, even the first resonance f_{R1} can support multiple fundamental frequencies, e.g. $f_{o,n(1)}$ at f_{R1}/n_1 or $f_{o,j(1)}$ at f_{R1}/j_1 with n and j being multiples of different supported fundamental frequencies . The two f_o which are both supported by a given f_{Rn}, are related by

$$f_{o,j(1)} = \frac{n}{j} * f_{o,n(1)} \qquad \text{or} \qquad f_{o2}/f_{o1} = \frac{n}{j} \qquad (1)$$

with f_{o1} and f_{o2} being the supported fundamentals respectively. The ratios of the integral numbers of the harmonics n and j define thus the resulting intervals.

Fig. 1 shows the case where both f_{R1} and f_{R2} support a fundamental frequency of 130 Hz. Besides the support at this frequency f_{o1} (130 Hz), f_{R1} supports the fundamental frequencies at $f_{o2} = 5/4 * f_{o1}$ which is a major 3rd and a mayor 6th ($f_{o2} = 5/3 * f_{o1}$) as well. f_{R2} will support a sharp whole tone (8/7), a perfect 4th (4/3), a minor 6th (8/5) and an octave (2/1).
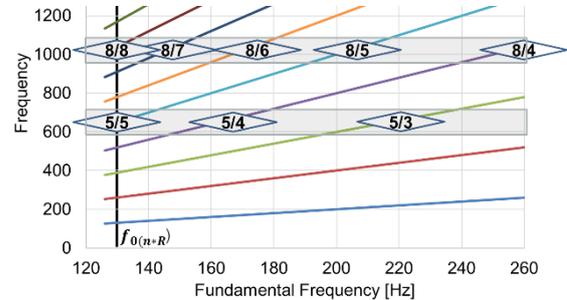
Figure 1: Schematic view of a pitch glide and the passage of harmonics across the f_{R1} and f_{R2} shown for the vowel [a:] of a male voice: fo_1 = 130 Hz, f_{R1}= 520 Hz and f_{R2} = 1170 Hz.

Fig. 2 gives a general view on supported intervals.

Figure 2: Formant-supported intervals based on a support of a given vocal tract resonance within the first 10 harmonics. **Black:** *intervals matching the diatonic system.* **Red:** *Intervals not found in traditional western music.*

Fig. 2 includes 10 pure intervals used in classical western music including the Tritonus while the minor second and a mayor 7th are missing. The intervals defined by the frequency relation of 9/8 are part of the traditional Indian Shrutis system while a lower minor 3rd and 4th are found e.g. in Arabian 24 scales. Thus, based on the purely physiological approach towards vocal tract ergonomic, different cultures may have chosen selected intervals to be »correct« in the sense that these intervals are equally supported by the human voice.

References

[1] Thomson, Academic Press, The Psychology of Music, 2013
[2] Ross *et al*, PNAS 104: 9852–9857, 2007
[3] Traser *et al*, JSLHR, 33: 2379-2393, 2017
[4] Ladefoged *et al*, 5th Conference on Speech, Communication and Technology, 1997
[5] Assmann *et al*, J. Ac. Soc. A 122: EL35-EL43, 2007

Measuring vocal-tract impedance at the lips: model, hypotheses and limits.

Timothée Maison[1]*, Fabrice Silva[1], Christophe Vergez[1], Nathalie Henrich Bernardoni [2]

[1]Aix-Marseille Univ., CNRS, Centrale Marseille, LMA, Marseille, France
[2]Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, Grenoble, France

Keywords: acoustic resonator ; radiation ; vocal tract impedance

Introduction

Non-invasive techniques (such as the so-called RAVE [1]) were developed to characterize vocal-tract acoustics using a broadband excitation and a microphone positioned closed to the lips. The measured pressure for an open-mouth condition, calibrated by a mouth-closed reference condition, provides estimates of the radiating vocal-tract resonance frequencies. Recently, exponential-sweep-based methods were reported to measure vocal-tract impedance [2, 3]. We propose to explicit the underlying hypotheses of such approaches, and to test their validity domain.

Methods

The measurement principle is based on the coupling between vocal tract and excitation tube, that can be assimilated as two waveguides. The pressure ratio given by the microphone at the lips can be expanded into

$$H = \frac{P^{measure}}{P^{reference}} = \frac{Z^{measure}}{Z^{reference}} \times \frac{U^{measure}}{U^{reference}} \ . (1)$$

The underlying assumptions of the method are: (i) the pressure at the lips for closed-mouth condition is supposed to be the same as the pressure of the capillary output, (ii) the acoustic current source for vocal-tract loading is ideal ($U^{measure} = U^{reference}$). Under the first assumption, the radiation coupling theory [4] leads to the impedance ratio:

$$\frac{Z^{measure}}{Z^{reference}} = \frac{Z^{VT}}{Z^{VT}+Z^{R}} \ . (2)$$

The first hypothesis is validated on a cylinder (L = 15 cm, d = 21 mm) by comparing measured impedances (with sensor [5]) of the left side of Eq. (2) to analytical computing of the right side of Eq. (2) (results not shown here). The second hypothesis (ii) is evaluated with a measurement at the lips (pressure ratio in Eq. (1)) — excitation tube and microphone on the same point at the cylinder inlet. The testbed is equivalent to RAVE [1] but using sweep excitation to characterize an open-closed cylinder as pseudo vocal-tract.

Finally, the robustness of the measure at the lips is explored by moving away horizontally excitation tube and microphone from the inlet, and computing frequency and quality factor ratio errors for the first three resonances.

Results

Figure 1 shows the comparison between a measurement at the lips (pressure ratio) and the numerical computation of Eq. (2). The relative errors on resonance frequencies are less than 1%, those on quality factors around 20%.
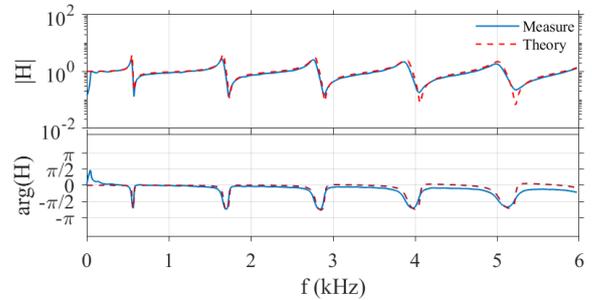
Figure 1: Comparison between experimental approach and numeric computing derived from the hypothesis for a cylinder.

Figure 2 highlights that the relative errors on first three resonance frequencies estimates are lesser than 1% as long as the distance between exciter/microphone and lips remains below 2 cm. Quality factors relative errors continuously increase with distance, reaching 80% at 2cm.
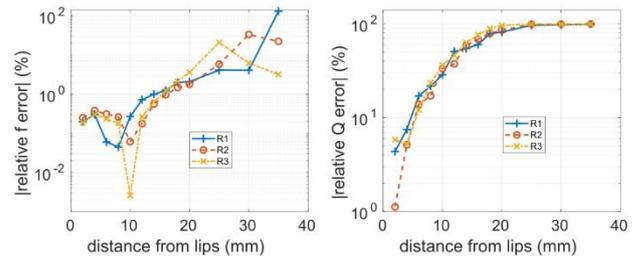
Figure 2: Relative error made on resonance frequencies and quality-factor measurements with cylinder inlet distance.

Discussion

This study focuses on a case with no additional voice source. The impedance ratio obtained with a measurement at the lips displays the resonances of the radiating vocal tract. The good correlation between measurements and numeric computation justifies the hypotheses: uniform pressure on lips plan and ideal source. If this method gives a robust access to resonance frequencies, with a possible distance gap to the lips, the quality-factor estimate suffers from the proximity between maxima and minima. Errors may also come from a long-range inefficient acoustic coupling. Further investigations are needed to improve quality-factor measurement, and to explore the case with voice source.

References

[1] Epps, Smith & Wolfe, Measure Sci. Tech., 8(10) :1112, 1997.
[2] Kob & Neuschaefer-Rube, Medical Eng. and Physics, 2002.
[3] Ahmadi & McLoughlin, IEEE ISCCSP, 2012.
[4] Chaigne & Kergomard, ed. Belin, 2008.
[5] Dalmont & Le Roux, JASA, 123(5) :3014-3014, 2008.

*maison@lma.cnrs-mrs.fr

Physical principle of using tubes for voice therapy methods demonstrated by experimental model of phonation.

Jaromír Horáček[1]*, Vojtěch Radolf[1], Vítězslav Bula[1], Anne-Maria Laukkanen [2]

[1]Institute of Thermomechanics of the Czech Academy of Sciences, Prague, Czech Republic
[2]Speech and Voice Research Laboratory, Faculty of Social Sciences, Tampere University, Tampere, Finland

Keywords: vocal tract acoustics; phonation into tubes; water resistance voice therapy

Introduction

Phonation through a tube either with the distal end in air or in water is widely used for voice training and therapy. This study investigates principles of both methods by combining results from computer modelling [1] and physical modelling [2].

Methods

The physical model consists of 1:1 scaled silicon vocal fold (VF) replica and a plexiglass tube representing the vocal tract (VT) for [u:] vowel. The VT was prolonged by the glass resonance tube (27 cm length, 7.8 mm inner diameter) with the other end in air or submerged 10 cm below water surface. Glottal area {GA) variation was registered with high speed camera. Formant frequencies were measured from the acoustic signal.

Results

The first formant frequency F1=333 Hz for [u:] decreased to F1=97 Hz for phonation through the tube into air and to F1=28 Hz for phonation through the tube into water. The computed frequencies are compared in Table 1 with the frequencies F1, the water bubbling frequency Fb and the fundamental frequencies *fo* measured in the model. Fig. 1 shows difference between vowel and tube phonation when studying the work done by airflow during the VFs self-sustained vibration. The loops constructed from transglottic pressure (Ptrans) and GA waveforms for tube phonation into air and water are similar, but for water not exactly periodic due to irregularities caused by bubbling.

Phonation type	f0 [Hz]	Fb [Hz]	F1 [Hz]
Vowel [u:]			
computation	/	/	333
experiment	110–113	/	316
Resonance tube in air			
computation	/	/	97
experiment	90–94	/	105
Resonance tube in water			
computation	/	/	28
experiment	75–86	19–24	26–28

Table 1: Computed acoustic resonance frequencies considering hard walls of the VT model, and the measured formant frequencies F1, for phonation on vowel [u:] and on [u:] with the VT prolonged by the resonance tube with the distal end in air and in water.
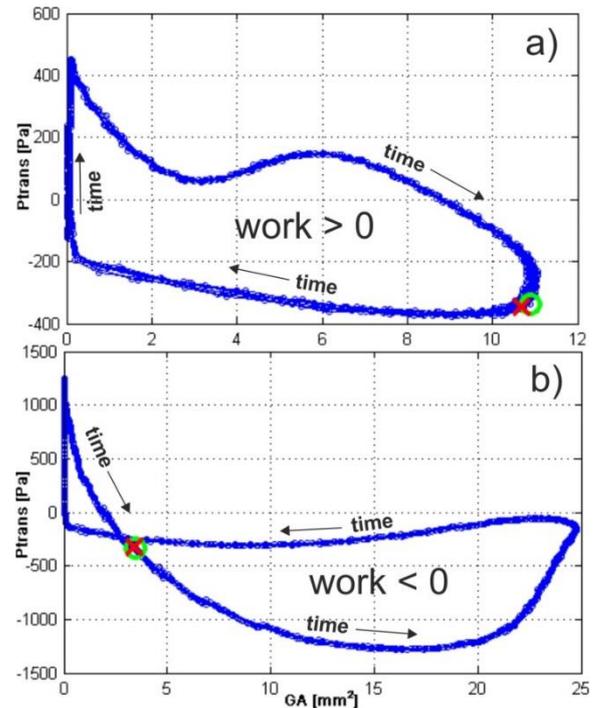
Figure 1: Examples of the loops Ptrans(t) vs. GA(t) measured during 9 periods of the VFs self-oscillation for phonation through: a) the VT, b) the resonance tube into air (Q=0.09 L/s, fo=94 Hz).

Discussion

The basic principle in vocal exercises with a resonance tube with the distal end in air or in water is the same. For phonation into air, *fo* excites the acoustic resonance at F1, and for phonation into water, water bubbling frequency excites a low frequency resonance which is caused by yielding walls of VT in humans. In both voice therapy methods, the part of the airflow energy required for phonation is substituted by the acoustic energy, utilizing the first acoustic resonance. Thus, less flow energy is needed to deliver the self-sustained vibration of the vocal folds.

Acknowledgements

References

[1] Horáček *et al*, Biomed Signal Proc and Control 37:39-49, 2017.
[2] Horáček *et al*, J Voice 33(4):490-496, 2019.

jaromirh@it.cas.cz

CT based FE modeling of the acoustic effects of nasality for vowels [a:] and [i:] in female voice

Tomas Vampola[1], Jaromír Horáček[2], *Anne-Maria Laukkanen [3]

[1]CTU in Prague, Dept. of Mechanics, Biomechanics and Mechatronics, Prague, Czech Republic
[2]Institute of Thermomechanics of the Czech Academy of Sciences, Prague, Czech Republic
[3]Speech and Voice Research Laboratory, Faculty of Social Sciences, Tampere University, Tampere, Finland

Keywords: voice biomechanics; resonances; singing; singer's formant

Introduction and aims

Classical singers use nasal consonants as 'resonance exercises', and aim at vibratory sensations at the nose and facial structures at all pitches sung. According to [1], a male singer used slight velopharyngeal opening for [a:] but not for [i:]. This was explained as a means to increase the prominence of the singer's formant during [a:] by decreasing the acoustic energy at F1 region. The nasal cavity may also serve as a narrowing for the main vocal tract, which might increase reactance over a wide range, thus supporting vocal fold vibration. This study investigates the effects of nasality on *fo* range.

Method

A novel, computerized tomography based model of the vocal tract and the nose from the same female volunteer was used. She sustained the vowels in speech range.

Results

Figure 1 shows reactance calculated over the frequency range 0-1.5 kHz for vowels [a:] and [i:] without and with nasality. With nasality, the lowest resonance frequencies R1-R2 increased and the reactance decreased throughout the range compared to non-nasalized phonation. Table 1 summarizes the lowest resonance frequencies for these two cases.

Discussion

The findings for R1 are in line with those observed for male speaker based on MRI [2] but differ for other resonances. Nasality seems to lower acoustic energy in general for the female. The only potentially beneficial effect of nasality in this participant might be that due to higher resonances, the vocal tract seems to be inertive at passaggio regions 300 Hz and 500 Hz and also at ca 1000 Hz which is near the high c

for sopranos. Also [3] reported that tenors in their study used velopharyngeal opening at passaggio.
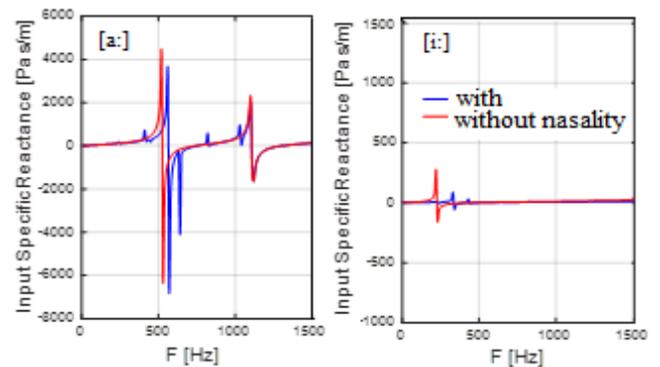
Figure 1: Reactance with and without nasality for vowel [a:] and [i:] in a female amateur singer.

[a:]	R1	R2	[i:]	R1	R2
without	599	1260	without	231	2209
with nasality	625	1265	with nasality	339	2233

Table 1: Lowest resonances for vowel [a:] and [i:] without and with nasality

Conclusions

Nasality may offer some assistance for phonation at passaggi. This requires further investigation.

References

[1] Sundberg J *et al*, J Voice 21(2): 127-137, 2007.
[2] Vampola T *et al*, Acta Acustica, 94:448-460, 2008.
[3] Birch P *et al, J Voice* 16(1):61-71, 20

*Anne-Maria.Laukkanen@tuni.fi

Prolonged Effect of TENS on Semi-occluded Vocal Tract in Teachers of Chillán

Jazmín Pérez[1], Francisca Carrasco[2], Rodrigo Fuenzalida[3], Alejandro Núñez[4]

[1]Fonoaudiología, Universidad Pedro de Valdivia, Chillán, Ñuble, Chile
[2]Fonoaudiología, Universidad Pedro de Valdivia, Chillán, Ñuble, Chile
[3]Fonoaudiología, Universidad Pedro de Valdivia, Chillán, Ñuble, Chile
[4]Fonoaudiología, Universidad Pedro de Valdivia, Chillán, Ñuble, Chile

Keywords : Transcutaneous electrical nerve stimulation (TENS)1; semi-occluded vocal tract (TVSO)2; acoustic parameters3; Voice4

Introduction: Teachers are the group of professionals who need to use their voice for prolonged periods of time, so they are more exposed to suffer vocal disorders[1]. TENS to treat swallowing and voice disorders is relatively new[2]. TENS uses percutaneous electrodes to transmit waveforms through the skin to stimulate large diameter nerve fibres. This stimulation triggers central inhibitory systems, which reduces fatigue, relaxes the muscles and causes better vascularization [3][4]. The TVSO exercises refer to a series of postures whose purpose is to lengthen and/or partially occlude the vocal tract, thus causing a change in the vibratory and resonant pattern of the vocal folds [5]6]. The reported benefits of the use of TVSO are greater economy in the production of the voice and changes in the pattern of vibration of the vocal folds. The objective of this research is to determine the prolonged effect of TENS on TVSO exercises in the voice of teachers from Chillán.

Methods: A quantitative, longitudinal study of descriptive-comparative level and quasi-experimental design was carried out. The sample consisted of 19 professors between 26 and 50 years old, with a minimum of 2 years of work experience and at least 30 teaching hours per week. The Teachers' voices were analysed with PRAAT software in the following parameters: Jitter, Shimmer, HNR and Maximum Phonation Time (MPT), pre-intervention, post-intervention (Re-ev.1), re-evaluation after a 2 month vocal rest (Re-ev.2) and re-evaluation again after 4 months of vocal load (Re-ev. 3). The intervention with TVSO consisted of: Control group: First session with tube phonation, Second session with ascending and descending glissandos. Third session with phonation in a tube with water. Fourth session with ascending and descending glissandos in a tube with water. Experimental group: IDEM to control and TENS.

Results: In the descriptive results, corresponding to the initial evaluation, it is observed that the average of the MPT values of the control group are better than the experimental group, but the standard deviation is more dispersed. The Jitter, Shimmer and HNR parameters of both groups appear normal, only local Shimmer is slightly above normal. In these parameters, the standard deviation of the control group is more dispersed in Jitter and Shimmer, while in HNR the dispersion is more accented in the exper-

imental group. It is noteworthy that in the second re-evaluation, which is after vocal rest, the indicators show a tendency to raise their disturbance levels in each vocal parameter. In this measurement the control group has higher perturbation levels than the experimental group. In re-evaluation 3 the levels decrease again in both groups, but the differences in means are more noticeable in the experimental group than in the control group. However, only significant intergroup differences are obtained in Local Shimmer of the control group and HNR in both groups.

Parámetros Acústicos	Evaluación X(SD)	Reev. 1 X(SD)	Reev. 2 X(SD)	Reev. 3 X(SD)	P
TMF Tens-TVSO	12,6 (3,28)	12,6(3,28)	13,44(5,22)	15,0 (5,22)	0,66
TMF TVSO	14,2 (3,61)	14,7((3,61)	14,20(3,61)	15,9 (10,07)	0,89
Jitter Tens-TVSO	0,6 (0,51)	0,3 (0,12)	0,44 (0,23)	0,3 (0,17)	0,24
Jitter TVSO	0,3 (0,11)	0,3 (0,19)	0,46 (0,99)	0,4 (0,17)	0,40
Shimmer TensTVSO	3,9 (4,04)	2,0 (0,35)	2,43 (0,94)	1,8 (0,64)	0,16
Shimmer TVSO	2,6 (0,99)	2,3 (1,02)	3,25 (0,91)	2,0 (0,68)	0,04
HNR Tens-TVSO	21,4 (4,56)	24,8(1,18)	21,8 (2,67)	24,8(2,88)	0,03
HNR TVSO	224 (2,53)	24,9(2,32)	19,9 (2,62)	23,2(3,32)	0,00

Table N°1: Averages(X) and standar deviation (SD) of acoustic parameters and intergroup p value with T of students for related samples. P <0.05 is considered. Control Group (TVSO). Experimental Group (TENS-TVSO).

Discussion: According to the results obtained in MPT, there were improvements in averages in both groups, which indicates improvement in glottic closure, similar to results obtained by Ras (2016)[7]. In the case of Jitter, Shimmer and HNR, the averages are improved in the final evaluation, however, only significant differences in HNR of both groups are observed. This differs from other authors who found no differences in any parameter (Mansuri, 2019; Guirró 2008) [8][9]. In TVSO there was a significant difference in Shimmer, but no better averages compared to TENS and TVSO, this differs compared to a study by Mansuri (2018)[10], who finds differences in Shimmer when applying TENS and Vocal Therapy. In conclusion, TENS can be used as a complementary therapy to improve TVSO results.

References

[1]Cobeta, I. *et al* Marge books,
[2] Freed M. *et al*, Respir Care, 46:466–474, 2001.
[3] Oliveira Santos *et al*, J. Voice 30(6), 769-e1, 2016
[4] Conde, M. *et al*, J Voice , *32*(3), 385-e17, 2018
[5] Vasquez K. *et al*, A. Universitarios*6*(2), 9-39, 2016
[6] Guzmán, M. *et al*, R.CEFAC, *14*(3), 471-480, 2012
[7] Ras,Y. *et al*,Egyptian J.Otolaryngology, 32(4), 322, 2016
[8] Mansuri, B. *et al*, Journal of Voice, *2019*
[9] Guirro, R. *et al*, R. Atualização Científica, *2018*
[10] Mansuri, B.*et al* , J. Voice, *2018*

*corresponding author e-mail

Effect of sleep and stress in voice functioning among college professors. A case study in a Colombian university

Andrés Carrillo-Gonzalez[1][*], Maryluz Camargo-Mendoza[2], Lady Catherine Cantor-Cutiva[1]

[1]Dept. of Collective Health, Universidad Nacional de Colombia, Bogotá, Colombia.
[2]Dept. of Speech and Language Pathology, Universidad Nacional de Colombia, Bogotá, Colombia.

Keywords: sleep quality; stress; voice acoustic analysis; college professors

Introduction

Sleep duration [1] and stress have been linked to general health outcomes. On the one hand, an inadequate sleep duration has negative effects on human functioning and may cause fatigue, affecting the voice performance [2]. On the other hand, stress can lead to adverse physiological, emotional, cognitive, or behavioral consequences [3]. An increased odd of reporting a voice disorder when stress is experienced has been previously pointed out [4]. However, to the best of the author's knowledge no research has still assessed the possible effect of sleep and stress in voice functioning in a college professor. Therefore, we performed a case study in a Colombian university with the aim of determining the relationship between sleep time and stress with three voice acoustic parameters (fundamental frequency, standard deviation of fundamental frequency, and standard deviation of vocal sound pressure level).

Methods

Case study of one college professor followed during 15-days. During the follow-up, daily self-reports from sleep quality and stress, along with daily voice records and sleep time measures were collected. We used Generalized Linear Models with Gamma Distribution to determine the relationship between sleep and stress with three voice acoustic parameters (fundamental frequency, standard deviation of fundamental frequency, and standard deviation of vocal sound pressure level).

Results

During a 15-days follow-up, fundamental frequency on connected speech increased very slightly when the participant reported longer sleep time (Figure 1) and decreased very slightly when the participant reported higher stress (B= 0.02 and -0.04, respectively) (Table 1). There was not a statistically significant association between standard deviation of fundamental frequency or standard deviation of sound pressure levels with stress or sleep.

Discussion

There is a very small, although significant, association between sleep time and self-reported stress with a reduction of fundamental frequency in connected speech. This reduction has been previously reported on people under stressful situations such as speaking in public [5].

From these results, we may conclude that hours of sleep and stress at work play an important role on voice production. Therefore, workplace vocal health promotion programs may want to include information on these aspects. However, since this is a case study, future research including bigger sample sizes are needed to confirm these results.
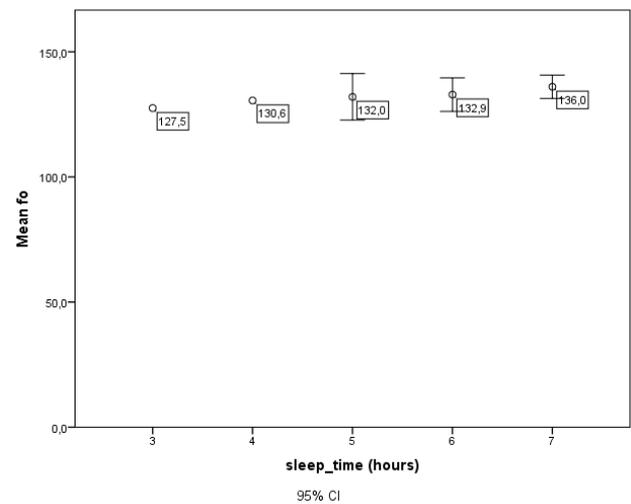

Figure 1. Relation between sleep time in hours and mean fundamental frequency

Parameter	UNIVARIATE MODEL			MULTIVARIATE MODEL		
	B	SE	p-value	B	SE	p-value
Sleep quality	0,00	0,02	0,95			
Sleep time	0,02	0,01	0,01	0,02	0,00	0,00
Stress	-0,03	0,01	0,01	-0,04	0,01	0,00

SE= Standard error

Table 1. Associations between sleep time, sleep quality and stress with fundamental frequency

Acknowledgements

References

[1] Cho *et al*, PLoS ONE, 12(8): e0182286, 2017.
[2] Bagnall *et al*, J Voice, 25(4):447-61, 2011.
[3] Giddens *et al*, J Voice, 27(3):390.e21-9, 2013.
[4] Kyriakou *et al*, J Voice, 32(5):643.e1-643.e9, 2018.
[5] Dietrich *et al*, J Speech Lang Hear Res, 55(3):973-87, 2012.

*acarrillog@unal.edu.co

*acarrillog@unal.edu.co

Identification of teachers at risk for phonotrauma using ambulatory monitoring of speaking fundamental frequency

Angélique Remacle[1]*, Laetitia de Chambourcy[2], Nathalie Lefèvre[3]

[1]Department of Speech Therapy, University of Liège, Liège, Belgium
[2]Faculty of Psychology and Educational Science, Université catholique de Louvain, Louvain-la-Neuve, Belgium
[3]Statistical Methodology and Computing Service, Université catholique de Louvain, Louvain-la-Neuve, Belgium

Keywords: Voice dosimetry; Vocal demand response; Risk factors; Occupational voice

Introduction

Excessive mechanical stress on the vocal folds – including deformation, oscillation, collision and acceleration [1] – can lead to tissue damage, that is, phonotrauma. Among professional voice users, excessive vibration patterns represent major risk factors for voice disorders [2].

This study evaluates several individual factors together to determine whether they can predict teachers' speaking fundamental frequency [f_0] and help to identify those who are most at risk of phonotrauma.

Methods

The f_0 of 87 teachers (66 women) without voice pathology was assessed every 200 ms during one workweek with the Ambulatory Phonation Monitor (KayPENTAX, Montvale, NJ), using an autocorrelation algorithm. The following individual factors were collected with questionnaires: gender, age, teaching experience, teaching level, tobacco consumption, gastro-esophageal problems, nonoccupational voice activity, voice education (i.e., having received weekly speaking and/or singing training for at least one year), past voice problems, and biopsychosocial impact of voice problems measured using the Voice Handicap Index [VHI [3]]). General linear mixed models were used to determine the effect of individual factors on teachers' f_0 tracked in real-life situations for a total of 431 days (4,479 hours).

Results

Speaking f_0 depended significantly on gender ($F(1, 74) = 132.1$, $p < .001$), teaching level ($F(3, 74) = 12.49$, $p < .001$), nonoccupational voice activity ($F(1, 74) = 4.34$, $p = .041$) and VHI score ($F(1, 74) = 8.95$, $p = .004$). Specifically, f_0 was higher in women, in individuals without nonoccupational voice activity, and in individuals with a higher VHI score (increase of 0.7 Hz for each additional point). For females, post hoc comparisons revealed a substantial impact of teaching level: university instructors had deeper voices than kindergarten ($p < .001$), elementary ($p < .001$), or secondary teachers ($p = .001$), and secondary teachers had deeper voices than kindergarten teachers ($p = .003$).

Individual factor	Females	Males
Gender	224.5	155.9
Without nonoccupational voice activity	229.3	144.0
With nonoccupational voice activity	219.2	140.6
Kindergarten (n=21, all female)	250.2	-
Elementary (n=20, all female)	236.9	-
Secondary (n=35, 20 females)	225.3	138.1
University (n=11, 5 females)	184.5	146.4

Table 1: Estimated marginal means for f_0 (Hz).

Discussion

The higher-pitched voices of kindergarten and elementary teachers may be due to the Lombard effect, the convergence effect, the characteristics of child-directed speech, and the cognitive and/or emotional load related to occupational stress [4]. On the other hand, lowering the pitch may be a strategy secondary and university teachers use to assert their authority. The lower f_0 of teachers who engage in nonprofessional voice activities may suggest acute inflammation or muscle fatigue due to voice overload. Prevention and early detection should be offered primarily to individuals at risk of phonotrauma due to higher f_0, namely females, and specifically those teaching at the kindergarten and elementary levels. In addition, self-assessment questionnaires such as the VHI could also help to detect individuals with potentially harmful f_0 patterns.

Acknowledgements

References

[1] Titze, J Voice, 8(2):99-105, 1994.
[2] Manfredi *et al.*, Logop Phon Vocol, 41(2):49-65, 2016.
[3] Woisard *et al.*, Rev Laryn Oto Rhin, 125(5):307-312, 2004.
[4] Van Puyvelde *et al.*, Front Psych, 9:1994, 2018.

* Angelique.Remacle@uliege.be

* Angelique.Remacle@uliege.be

Costs associated to voice disorders in Colombian telemarketers

Harold Zamir Taborda-Osorio[1], Lady Catherine Cantor-Cutiva[1*]

[1]Speech Language Pathology Program, Universidad Manuela Beltrán, Bogotá D.C, Colombia

Keywords: Health care Costs, Voice Disorders, Telemarketers

Introduction

Occupational voice users, such as teachers and telemarketers have higher likelihood of developing voice disorders. According to Cantor-Cutiva, Vogel & Burdorf [1], teachers have a high occurrence of voice problems with a current prevalence ranging from 9% to 37%, and an annual prevalence that can vary between 15% and 80%. In addition, teachers are professionals who show a higher rate of voice problems (dysphonia) according to some studies in the field [2,3,4]. In the case of telemarketers, previous studies have reported that they are frequently exposed to different risk factors such as inadequate ergonomic aspects, sudden temperature changes, rooms without acoustic treatment, dust, stressful workplaces, need for more breaks, difficulties in the relationship with bosses, among others [5,6]. These workers, on average, suffer 64% of dry throat, 33% neck and neck pain, 31% hoarseness 26% voice failures and 22% vocal fatigue [7]. Therefore, we performed this retrospective study with the aim to determine the costs related to voice disorders in Colombian Telemarketers.

Methods

Retrospective study on information related to the occurrence and associated costs of voice disorders in telemarketers. The analysis will focus on the occurrence of voice disorders and economic consequences, such as absenteeism and medical consumption due to voice disorders in the last 5 years. The research will carried out in a multinational company that provides services internationally and serves different companies in the health, transportation, energy, food, service and entertainment sectors in Colombia.

Expected results

Occurrence of voice disorders in telemarketers may generate economic losses to the company, similar situation found in teachers[1]. Problems of absenteeism in the educational sector and contribute to addressing the weaknesses of economic and administrative approaches to the phenomenon[8], features found in the work of telemarketers of this study.

References

[1] Cantor-Cutiva LC, Vogel I & Burdorf A. Voice disorders in teachers and their associations with work-related factors: a systematic review. J. Commun. Disord, 46:143–155. 2013.

[2] Behlau M, Zambom F; Guerrieri AC & Roy N. Epidemiology of voice disorders in teachers and nonteachers in Brazil: prevalence and adverse effects. J Voice, 26: 665.e9-18. 2012.

[3] Van Houte E, Claeys S, Wuyts F & Van Lierde K. The impact of voice disorders among teachers: vocal complaints, treatment-seeking behavior, knowledge of vocal care, and voice-related absenteeism. J Voice, 25: 570-575. 2011.

[4] Simões-Zenari M, Bitar ML & Nemr NK. The effect of noise on the voice of preschool institution educators. Rev Saude Publica, 46 : 657- 664. 2012.

[5] Fererira, L; Giannini, S, Latorre, M & Zenari.. Vocal disorders related to work: proposing a tool to evaluate teachers. Distúrb da comunicação; 19: 127-37. 2007.

[6] Piwowarczyk TC, Oliveira G, Lourenço L & Behlau M. Vocal symptoms, voice activity, and participation profile and professional performance of call center operators. J Voice, 26: 194-200.

[7] Amorim GO, Bommarito S, Kanashiro CA & Chiari BM. Comportamento vocal de teleoperadores pré e pós-jornada de trabalho. J Soc Bras Fonoaudiol, 23: 170-176. 2011.

[8] de Medeiros AM, Assunção AÁ & Barreto SM. Absenteeism due to voice disorders in female teachers: a public health problem. Int Arch Occup Environ Health., 85: 853-864.

Tuning MRI-based vocal tracts to modify formants
in the three-dimensional finite element production of vowels

Marc Arnela*, Oriol Guasch, Arnau Pont

GTM Grup de recerca en Tecnologies Mèdia, La Salle, Universitat Ramon Llull, Barcelona, Catalonia, Spain

Keywords: acoustic sensitivity functions; finite element method; MRI-based vocal tracts; vowels

Introduction

Accurate three-dimensional (3D) vocal tract (VT) geometries from Magnetic Resonance Imaging (MRI) (see e.g., [1]) can be used for the numerical simulation of sounds like vowels [2], diphthongs [3] and sibilants [4]. These VT geometries may need to be modified to match the formant frequencies from natural recordings, or to simply generate new sounds such as in singing voice. Vocal tract variations can be driven following the tuning strategy in [5], in which acoustic sensitivity functions were applied to 1D VTs described by an area function. In this work we extend the methodology in [5] to modify 3D MRI-based VTs by first converting them to 1D, then proceed to tuning and finally reverting back to 3D geometries.

Methods

The MRI-based VT geometries for static vowels sounds in [1] were selected for this work. First, they were adapted for our purposes, removing the face, lips, and subglottal tract. Second, they were discretized in a set of cross-sections along the vocal tract midline, as typically done to extract 1D area functions. However, not only the area of each cross-section was computed, but also their shape and location through the vocal tract midline. On the other hand, the mixed wave equation characterizing the 3D vowel generation problem was converted into a 1D Webster equation in the frequency domain, with complex wavenumber. Its eigenvalues and eigenvectors were obtained by means of the finite element method (FEM). A modal perturbation analysis followed which resulted in the same sensitivity functions of [5], but without the need to resort to non-linear radiation pressure. The sensitivity functions were then used in the iterative algorithm of [5], which induces variations in the 1D vocal tract area functions until formants reach prescribed target values. A 3D vocal tract geometry was then reconstructed from the final 1D area function, making use of the stored original shape and location of each VT cross-section.

FEM simulations were performed to analyze the vocal tract acoustics of the original and modified vocal tract geometries, using the in-house code outlined in [3]. The vocal tract transfer function $H(f)$ was computed for each configuration. Here, $H(f) = P_o(f)/Q_i(f)$, with $P_o(f)$ and $Q_i(f)$ respectively standing for the Fourier transform of the acoustic pressure at the mouth and the input volume velocity at the glottis.

Results

Figure 1 shows some results in which the MRI-based vocal tract of vowel [ɑ] is modified to produce a cluster of the third, fourth and fifth formants, as typically observed in singing voice.
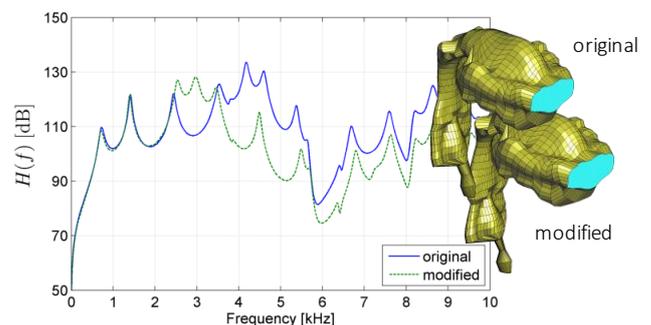

Figure 1: Vocal tract transfer function $H(f)$ of vowel [a] for the original and modified 3D vocal tract geometries.

Conclusions

In this work, we have presented a methodology to tune the formants of 3D MRI-based vocal tracts, which has been applied to the FEM generation of vowel sounds. The swapping between 3D and 1D allows one to modify planar mode resonances, while keeping the higher frequency content of voice dictated by the 3D VT geometry.

Acknowledgements

References

[1] D. Aalto *et al*, "Large scale data acquisition of simultaneous MRI and speech," Appl. Acoust., 83:64–75, 2014.
[2] M. Arnela *et al*, "Influence of vocal tract geometry simplifications on the numerical simulation of vowel sounds," J. Acoust. Soc. Am., 140(3):1707–1718, 2016.
[3] M. Arnela *et al*, "MRI-based vocal tract representations for the three-dimensional finite element synthesis of diphthongs", IEEE/ACM Trans. Audio Speech. Lang. Process., 27(12):2173-2182, 2019.
[4] A. Pont *et al*, "Computational aeroacoustics to identify sound sources in the generation of sibilant /s/". Int. J. Numer. Meth. Biomed. Eng., 35(1):e3153, 2019.
[5] B. H. Story, Technique for "tuning" vocal tract area functions based on acoustic sensitivity functions, J. Acoustic. Soc. Am., 119(2):715-718, 2006.

*marc.arnela@salle.url.edu

Towards an open-access database of 3D shapes of the vocal tract and their aero-acoustical properties

Mario Fleischer[1]*, Peter Birkholz[2]

[1]Department of Audiology and Phoniatrics, Charité – Universitätsmedizin Berlin, Berlin, Germany
[2]Institute of Acoustics and Speech Communication, Technische Universität Dresden, Dresden, Germany

Keywords: Vocal tract, Aero-Acoustics, Numerical Modelling, Speech dataset

Introduction

For a comprehensive and profound understanding of the origin and transmission of acoustic energy in the human vocal system, representative three-dimensional geometric descriptions of the vocal tract (VT) shape are essential.

Methods

We determined the air-filled cavities of the VT of one male and one female speaker by using 3D Magnetic Resonance Imaging. The subjects were asked to produce 22 German speech sounds each (16 vowels and 6 consonants). For all resulting 44 models, the teeth of the subjects were added to the raw data before segmentation.

Based on the processed 3D images, we built a complete set of volume meshes needed for numerical analyses and half-shell 3D-printable models (Figure 1).

All vocal tract models were acoustically characterized in terms of their transfer functions (0-10 kHz) that were determined both numerically with the 3D Finite Element Method, and experimentally by means of an external source excitation paradigm [1]. Furthermore, for the analyses of fricatives and in terms of human recognition of the acoustic outcome, all 3D-printed models were excited by using an artificial flow source at the glottis [2].

Results and Discussion

We found an excellent agreement (with an average difference of below 1%) of the experimentally and numerically determined frequency values of the first three vocal tract resonances (f_{R1}, f_{R2} & f_{R3}). Moreover, the first two resonances of the vowels were plausibly arranged in the f_{R2}-f_{R1}-plane and clearly separated from each other. That indicates that all vowels were produced differently from each other, despite the non-ideal recording conditions in the MRI scanner and the need to artificially sustain the vowels for 12 seconds.

We plan to make all raw data, models and results freely available on an open-access platform to the scientific community. The actual state of this project, including findings of flow-induced sound experiments, will be presented and discussed.
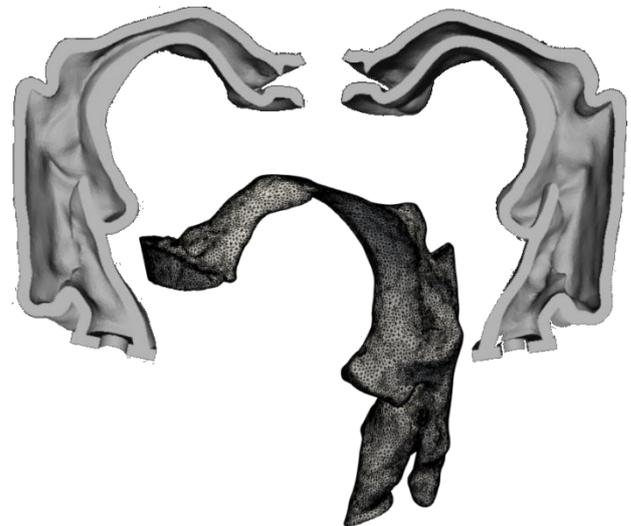

Figure 1: Example for a 3D-printable vocal tract model and the associated Finite Element Mesh

References

[1] Fleischer *et al*, PloS ONE, 13(3):e0193708, 2018.
[2] Birkholz *et al*, J Acoust Soc Am, 146(1):223-232, 2019.

mario.fleischer@charite.de

Monkey vocal tracts are not so "speech ready"

Frédéric Berthommier[1]

[1] Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, 38000 Grenoble, France

Keywords : monkey vocalizations, vocal tract states, proto-vowels

Introduction

Monkey vocalizations are important for studying evolution of speech. During about 50 years, this was considered as blocked until *Homo Sapiens* since the anatomy of monkey and "other nonhuman primate" is not set for producing upper vocal tract formant modulations [1]. Two recent publications [2,3] claimed against this theory that monkey vocal tract is "speech ready" and that a high larynx combined with a flat tongue are not a handicap. The main message is that only the brain had to evolve to control an existing tongue musculature able to shape constrictions and cavities as human. It has been shown that (1) baboons can produce some vowel qualities [3] (2) some vocal tract shapes compatible with vowel production are present on sagittal X-Ray views of the macaque [2]. This has been completed by [4] with a collection of published spectrograms.

Results

Considering all of these results, we observe that vowel quality /i/ is never produced, as well as the vowel quality /a/ which shape is retrieved on sagittal X-Ray views only. We develop the hypothesis that it is the consequence of the lack of tongue rounding necessary to make various constriction locus along the palate and the back of the (too small) pharyngeal cavity. First, in order to analyze the articulatory potential of monkeys, we have built a set of tube-models comparable to sagittal views (Figure 1). We see that for /u/ only a small pharyngeal cavity is remaining after tongue retraction (state 1). The main vowel qualities observed and compiled by [3,4] are close to the /u/ and in this framework they are well explained by mouth opening and tongue protrusion derived from this configuration. An exception is the baboon's wahoo vocalization and a second specific model has been developed to gauge the articulatory potential of animal in comparison with human. This model shows that /wa/ can be produced without constriction, but that the /wa/-/hoo/ transition is different for man and baboon because with a flat tongue, the constriction place cannot be moved from forward to backward. This is another direct consequence of monkey's anatomy.

Conclusion

With a flat tongue and a high larynx, an archaic articulatory potential is present allowing the production of proto-vowels. Nevertheless, the monkey cannot produce every articulatory movement necessary for speech production and the vocal tract anatomy had to evolve further.
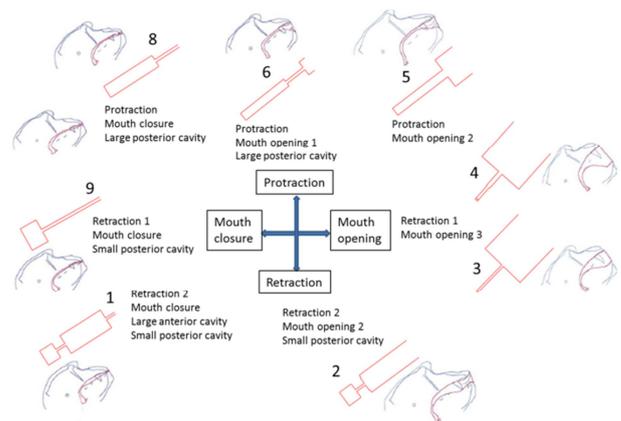

Figure 1: Direct anatomical comparisons between tube models and sagittal drawings provided by [2], concerning tongue-palate constriction, mouth aperture and the size of the back cavity. The articulatory states (1-9) are identified with simple rules: (a) the tongue is protracted or retracted with 2 levels (see [5]) (b) the mouth is the anterior part of the oral cavity which includes the tongue tip and the lips, and it is close or open with 3 levels (c) when the tongue is retracted, a back constriction forms a small posterior cavity (1,2,9) or a thin uniform section (3) (d) the tongue tip is entangled with the tongue and when this is retracted at level 2, the tip is also retracted to make a large anterior cavity with close lips (1) or a uniform section with open lips (2). Finally these variations are well described with two parameters for mouth and tongue well controlling F1/F2 variations. This leads to 3 extreme states having high constrictions ratio and comparable to those of the human vowel qualities /u/ (1), /i/ (8) and /a/ (3,4). This is also compatible with the basic IPA description of the human vowel chart in which F1 variation follow jaw opening/closing and F2 variation follows the front/back tongue movement.

References

[1] Lieberman, P. et *al*, Science, 164(3884):1185–1187, 1969.
[2] Fitch, W. T. et *al*, Science Advances, 2(12), e1600723, 2016.
[3] Boë, L.-J. et *al*, Plos one, 12(1), e0169321, 2017.
[4] Boë, L.-J. et *al*, Science Advances, 5(12), eaaw3916n, 2019.
[5] Hiiemae, K.M. et *al*, Arch. of Oral Biol., 40(3):229-246, 1995.

*frederic.berthommier@gipsa-lab.grenoble-inp.fr

The role of between- versus within-speaker acoustic variability in vocal identity perception

Jody Kreiman[1]*, Yoonjeong Lee[2]

[1]Departments of Head and Neck Surgery and Linguistics, UCLA, Los Angeles, CA, USA
[2]Department of Head and Neck Surgery, UCLA, Los Angeles, CA, USA

Keywords: speaker recognition; acoustic variability; prototypes; personal identity

Introduction

Our recent studies [1] show that acoustic spaces characterizing within- and between-speaker variability in voice quality have similar structures, with a few features (acoustic variability and formant dispersion) that are prominent for all speakers combined with idiosyncratic features characterizing individual talkers. These findings suggest that voice discrimination ("telling voices apart" [2]) should be based on shared acoustic features, while determining which samples come from a single talker ("telling voices together" [2]) should depend on knowledge of each individual's vocal idiosyncrasies. Based on prototype models of voice perception, we hypothesized that errors in telling voices apart would be strongly predictable from distances among those few shared features in the group acoustic space, but that errors in telling voices together would not be associated with these distances. This experiment tested that hypothesis.

Methods

Recordings of one side of unscripted telephone conversations from 49 female speakers were drawn from the UCLA Speaker Variability Database [3]. Based on the acoustic structure of individual and group voice spaces for these speakers, 8 were identified as acoustically prototypical (P1, P2, P3, P4, P5) and aprototypical (A1, A2, A3) for this population of talkers. Fourteen brief samples were excerpted from each conversation for use in this study (phrases or sentences; duration: M = 1.6 s, SD = 0.5 s). Samples were screened to ensure they did not contain identifying content or unusual features (repeated phrases, laughs, speech errors, etc.) and were normalized for intensity.

Stimuli were sorted into 3 sets of 4 trials, such that across sets listeners heard unique pairs of speakers, and across trials each listener heard every speaker, but no listener heard any speaker more than once. Each trial included the 14 speech samples from 2 talkers, represented as randomly colored and shaped icons in a PowerPoint slide (Figure 1). Each set of 4 trials was judged by 10 listeners (30 total).

Listeners were instructed to listen to the voices, and then to drag them into piles on the screen so that each pile corresponded to a single speaker. They were told that there could be different numbers of voices in the piles,

and that they could put as many voices as they liked in each pile, depending on how many speakers they thought they heard. (They were not told that there were in fact only 2 speakers.) They were allowed to listen to the samples as often as needed, in any order. The complete experiment lasted about 1 hour.
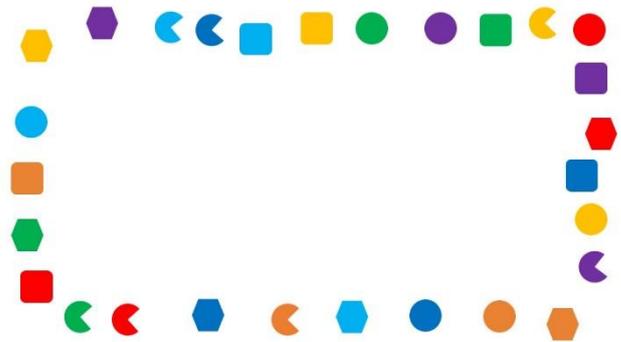

Figure 1: The testing interface. Each icon represents a single voice sample. Listeners used a mouse to drag these icons into piles corresponding to perceived speaker identities.

Preliminary Results and Discussion

Subject testing is currently underway. Most listeners in pilot studies found the task challenging. Results to date are consistent with our hypotheses. Overall, listeners showed fewer errors in the "telling voices apart" task than in the "telling voices together" task. Prototypicality in voices played an important role in voice discrimination. Separate analyses will shed light on the features and strategies involved in these kinds of judgments, with particular attention to features that help listeners tell voices together. Implications for prototype models of voice perception will be discussed.

Acknowledgements

References

[1]Lee *et al.*, J Acoust Soc Am, 146:1568-1579, 2019.
[2]Lavan *et al*., Psychon Bull Rev, 36:90-102, 2019.
[3]Keating *et al.*, Proc ICPhS, 2019

*jkreiman@ucla.edu

Vocal strategies to signal biological fitness in public speaking: a study on the effects of aging in American English charismatic speech

Rosario Signorello[1]*, Didier Demolin[1]

[1]Laboratoire de Phonétique et Phonologie (UMR7018, CNRS - Sorbonne Nouvelle), Paris, France

Keywords: voice quality; public speaking; aging and disordered voice; biological fitness

Introduction

Voice signal speaker's biological fitness through language-based learned ways to convey it [1, 2]. Speakers involved in public speaking perform voice quality manipulations to convey charisma in order to influence others emotionally [1]. Charismatic voice defines the ensemble of vocal quality patterns used by public speakers with the intent to persuade the audience: it conveys personality traits and arouse emotional states in listeners. Listeners exposed to public speaking rely on charismatic voice to assess effective leadership [1].

Perceived vocal quality features could then lead to misunderstanding speakers' biological fitness and, in turn, to affect the effectiveness of leadership. It is, for example, difficult to guess speakers chronological age from voice, as aging voice is difficult to assess both in its acoustic and perceptual correlates. A previous study [2] had indicated that perceived vocal disorder, which is the acoustical and perceptual variation of an expected normal quality of the sound of the voice, was the only way that listeners have to precisely assess how much someone has aged. Vocal disorder seemed to form the bridge between charismatic voice and aging in voice: political leaders that had experienced vocal disorder suffered from an impact to their charisma and political career since listeners were very good at rating speakers' chronological age, if the individual has vocal disorder [3].

The goal of this study is to investigate the vocal strategies that American English charismatic leaders use to cope with the acoustical and perceptual effects of chronological aging as possible sign of lack of biological fitness in vocal performance while engaging in public speaking.

Methods

Voice acoustic data were collected for American English charismatic leaders Donald Trump (68–69 years old at the time of the recording), Bernie Sanders (74 y.o.), and Barack Obama (51–53 y.o.). Data were recorded in different periods of time (during the 2012, 2016, and 2020 United States presidential primaries) and while engaging in three different contexts of communication: a monologue addressed to followers (MON), a monologue addressed to other politicians (CON), and an informal interview during which the speaker addressed one listener only (INT). Longitudinal and cross-sectional analyses were performed on fundamental frequency (f0) and sound pressure level (SPL) measured from /a/ vowels.

Results

Preliminary cross-sectional analyses of average f0 and SPL show that speakers Obama ($t(149)=6.27$, $p<.001$, $r=.45$) and Trump ($t(952)=16.44$, $p<.001$, $r=.47$) present stronger positive correlations between the two acoustic parameters than speaker Sanders ($t(113)=2.94$, $p=.003$, $r=.26$) while addressing a large audience (context MON). Additionally, speakers Obama ($t(121)=5.52$, $p<.001$, $r=.44$) and Trump ($t(157)=5.03$, $p<.001$, $r=.37$) presents positive correlation between average f0 and SPL while addressing other politicians (CON context), differently than speaker Sanders, who presents non-significant correlation ($p>.05$). On the contrary, speaker Sanders ($t(29)=7.89$, $p<.001$, $r=.82$) presents stronger positive correlations between f0 and SPL than speakers Obama ($t(161)=4.97$, $p<.001$, $r=.36$) and Trump ($t(172)=8.36$, $p<.001$, $r=.53$).

Discussion

This study investigates aging voice as possible sign of lack of biological fitness in public speaking. On one hand, correlated f0 and SPL in one-to-many charismatic speech (MON and CON) could be a signal of low chronological age (even in discrepancy with actual age) and high biological fitness. This vocal behavior would be a positive predictor for effective charismatic leadership [1]. On the other hand, weaker or negative correlation between f0 and SPL in one-to-may charismatic speech (MON and CON) and higher or positive correlation in one-to-one speech (INT) could show advanced chronological age (even in accord with actual age) and low biological fitness. This vocal behavior would be a negative predictor for effective charismatic leadership [1].

References

[1] Signorello et al, J Voice, 45:1-10, 2018.
[2] Kreiman and Sidtis, Foundations of Voice Studies, 2011.
[3] Signorello and Demolin, JASA, 142: 2641-2641, 2017.

rosario.signorello@sorbonne-nouvelle.fr

Automatic fundamental frequency characterization
of premature newborns' cries in Neonatal Intensive Care Unit

Bertille MET-MONTOT[1]*, Sandie CABON[1], Patrick PLADYS[1], Guy CARRAULT[1], Fabienne POREE[1]

[1]Univ Rennes, CHU Rennes, Inserm, LTSI – UMR 1099, F-35000 Rennes, France

Keywords: prematurity; monitoring; cry analysis; frequency estimation

Introduction

Newborn cries have been widely studied in the literature and have been proven to be relevant for the detection of pathologies and for the assessment of the neurobehavioral development [1]. However, to date, solutions to automatize this analysis in care routine remain to be proposed. The European project Digi-NewB aims to address this point by developing a new generation of monitoring system in Neonatal Intensive Care Unit (NICU), that will combine signals from different sources, including electro-physiological, video and audio data.

This work deals with sound aspects and has for first objective to automatically extract and characterize newborn cries in long recordings and in a noisy acoustic environment (e.g., adult voices, alarms). A focus is made on the automatic estimation of the fundamental frequency and its evolution along a cry (also called melody), using an image processing method applied to the spectrogram.

Methods

The developed automatic processing chain is schematized in Figure 1 and is divided into six main steps that are described hereafter:
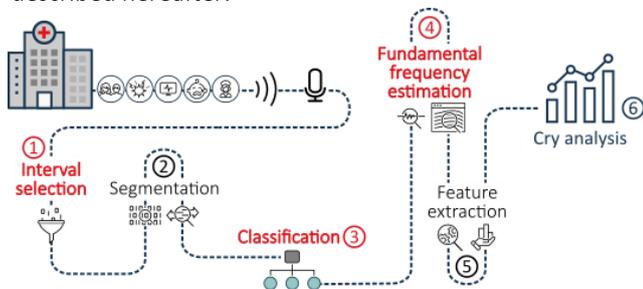

Figure 1: Diagram of the overall processing chain of the study. The steps in red represent the contributions made by our team.

1. Selection of intervals of interest to analyze only recordings where newborns cry. This step is based on the energy level of the soundtracks.
2. Extraction of all sound events from the recordings by a short-term energy thresholding described by [2].
3. Separation of cry segments and noisy ones using a classification model (KNN) [3].
4. Detection and tracking of the fundamental frequency using an image processing method applied to the spectrogram.
5. Extraction of the most suitable temporal and frequential features (e.g., duration, fundamental frequency, melody).
6. Analysis of the cry characteristics.

This contribution focuses on the estimation of the fundamental frequency and its evolution over time. According to the literature, this estimation is usually done in a fixed and manually chosen frequency band, that implies two types of limitations: i) the poor estimation of hyperphonic cries (whose fundamental frequency is higher); ii) the possible jumps in the fundamental frequency estimation when high-energy formants are included in the study frequency band.

The method developed here is based on:

1. Computation of the spectrogram of a cry;
2. Automatic detection of the frequency sub-band containing the fundamental frequency;
3. Detection of all the contours present in this band;
4. Selection of the longest contour as the main component of the melody of the cry.

Results and Discussion

From the first three steps of the processing chain, a total of 5016 segments were automatically extracted from 18 newborns and 240 hours of recordings. Figure 2 shows the results of the frequency estimation method applied on two different types of cries. It can be observed that the melody is well tracked by the algorithm in both cases.
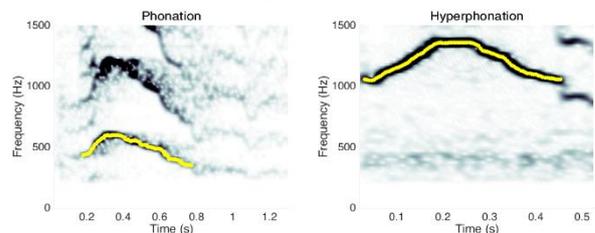

Figure 2: Fundamental frequency estimation for two type of cries: a phonation cry (left) and an hyperphonation cry (right).

This method will now allow an automatic and accurate characterization of the cries. To this end, we will process the thousands of cries that were extracted in order to identify the most relevant features for the evaluation of the neurobehavioral development of premature newborns.

Acknowledgements

References

[1] S. Cabon *et al*, Physiol. Meas. 40:02TR02, 2019.
[2] S. Orlandi *et al*, J BSPC, 8:799-810, 2013.
[3] S. Cabon, PhD thesis Univ Rennes 1, 2019.

How changing the vocal fold vertical stiffness gradient changes vocal fold vibration

Sid Khosla , Liran Oren , Charles Farbos de Luzan

Department of Otolaryngology, University of Cincinnati, Cincinnati, Ohio, USA

Keywords: vocal fold elasticity, phonation, mucosal wave

Introduction: During vocal fold vibration, a coronal section of the vocal folds is convergent during opening and divergent during closing; we will refer to this vibratory pattern as the vertical mucosal wave (VMW). We also   know that , as displacement   increases, the superior aspect of the folds becomes less stiff than the inferior aspect; we will refer to the change in stiffness as the vertical stiffness gradient (VSG).   These two facts led us to hypothesize that reduction of the vertical stiffness gradient will reduce the vertical mucosal wave.

Methods:4 excised canine larynges were used. As previously described, particle imaging velocimetry of the intraglottal velocity fields was done at low and high subglottal pressures in the mid coronal plane .This was considered the baseline, and using the indentation method, stress strain curves were taken for the superior and inferior 1 mm of the fold in the mid membranous plane. Radiesse with calcium hydroxyappetite (CaHa) crystals was then injected into the very superior aspect of the fold. New stress strain curves were then obtained. PIV was then taken and intraglottal velocity fields were obtained.

Results: Injecting   the CaHa reduced the inferior-superior stiffness gradient by over 75% in all larynges. The maximal divergence angle was also markedly reduced resulting in minimal VMW.

Discussion: Marked reduction of the vertical stiffness gradient also significantly reduces the vertical mucosal wave. Clinical implications will be discussed in addition to possible hypotheses for reduced VMW.

corresponding author e-mail

Measurements of phonatory power flows and efficiencies in a human airway phantom

Michael Krane*, Paul Trzcinski, Gage Walters, Zachary Yoas

[1]Applied Research Laboratory, Penn State University, State College, Pennsylvania, United States of America

Keywords: voice efficiency; Acoustics; Aerodynamics and kinematics; Modeling

Introduction

We present empirical characterization of phonation power flows and efficiencies for three distinct regions of a human airway phantom: the larynx, the vocal tract, and the combination of larynx and vocal tract. A control volume power flow formulation [1], which identified inputs, outputs and losses in terms of joint statistics between volume flow and pressure, guided the measurements. New results include refinements to the efficiency measures, showing an explicit relation to laryngeal resistance, and showing how previously presented measures of efficiency [] are incorporated into the current formulation. The measurements presented here are used to estimate the terms in the energy equation for the larynx, the vocal tract and the system composed of their combination. From these estimates the efficiencies are also estimated.

Methods

Refinements of a previous mechanical energy equation analysis [1] are presented. These refinements consist of cycle-averaging and incorporating features such as mean laryngeal flow resistance. In addition the importance of volume flow-pressure covariances is highlighted. The terms of the resulting averaged equations represent power flows to or form the region in question, the direction of the power flow being given by its sign. The terms in this formulation were estimated from measurements. In addition, the ratio of average output to average input power flow was taken as the definition of efficiency for each region.
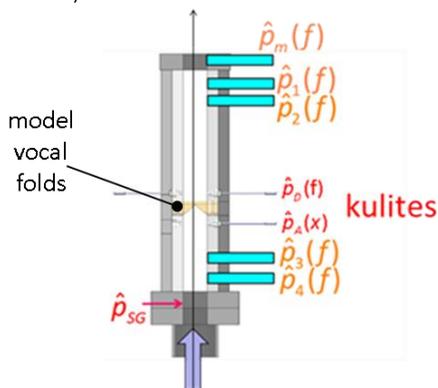

Figure 1: Schematic of Penn State Upper Airway model. Locations of pressure measurements shown − 5 microphones to measure sound pressure and 2 kulite XCS-093 pressure transducers to measure transglottal pressure difference. A high-speed camera (not shown) is used to capture vocal fold motion.

Measurements were performed in the Penn State Upper Airway phantom [1-3], shown in Figure 1. Air was supplied from a compressed air source through a 7.62cm diameter tygon tube. Upstream of this tube a thermal flow meter was installed to measure average air mass flow. The tygon tube connected to the airway phantom, which has a 2.54cm square cross section, with a trachea section of length 10cm, a vocal tract section of length 17.3, and a larynx section containing multilayer rubber molded vocal folds with a swept-ellipse section. Pressure measurements were performed in the airway phantom: microphones at the mouth opening, a pair of microphones near the mouth opening, and another microphone pair near the trachea entrance. In addition, two Kulite pressure transducers were installed, one just upstream and the other just downstream of the model vocal folds, to measure the transglottal pressure difference. The microphone pairs were used to estimate the right- and left-running wave pressures and velocities associated with the acoustic field. These were then used to estimate the relevant variances and covariances that appear in the power flows and efficiencies from the auto- and cross-spectra estimated from the microphone pair measurements.

Results

Measurements were conducted for two vocal fold models that exhibit different vibration patterns. The first of these is a pair of identical three-layer rubber vocal folds, and exhibits a periodic and the second is composed of model vocal fold of this same type and one composed of a single layer. This setup results in two folds with different vertical stiffness, resulting in a quasiperiodic vibration pattern with reduced sound production.

The pressure measurements are used to estimate power flows and efficiencies as described above. A particular result concerns the importance of pressure-volume flow covariances at the laryngeal inlet and outlet.

Acknowledgements

References

[1] Krane et al, Proc. of 11th ICVPB, E. Lansing, MI, 2018.

[2] Schutte, H. The efficiency of voice production. PhD diss. 1980.

[3] Isshiki, N. Vocal fold physiology, 193-203, 1981.

[4] Titze, et al., J. Voice, 30(4) 398-406, 2015.

mhk5@arl.psu.edu

[5] Campo, L. MS Thesis, Penn State Univ., 2012.

[6] McPhail, et al *JASA* 146(2) 1230-1238, 2019.

*mhk5@arl.psu.edu

Control volume analysis of glottal jet dynamics using time resolved pressure and velocity field measurements in a scaled up vocal fold model

Timothy Wei[1]*, Hunter Ringenberg[1], Dylan Rogers[1], Nathaniel Wei[1], Lucy Zhang[2], Michael Krane[3]

[1]Mechanical & Materials Eng'g., University of Nebraska-Lincoln, Lincoln, NE, USA
[2] Mechanical, Aerospace & Nuclear Eng'g., Rensselaer Polytechnic Institute, Troy, NY, USA
[3]Applied Research Laboratory, Pennsylvania State University, State College, PA, USA
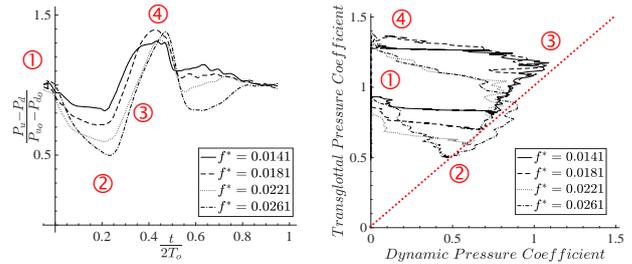
Figure 1: Transglottal pressure (TGP) vs time at Re = 7200 for the four frequencies. ① Start of cycle, ② TGP is a minimum, ③ Dynamic pressure is a maximum, ④ TGP is a maximum.

Figure 2: Phase plot of transglottal pressure and dynamic pressure for the four different frequencies at Re = 7200. Red circled numbers correspond to those shown in

Keywords: glottal jet dynamics; pressure-velocity correlations; control volume analysis; aeroacoustic sources

Introduction

This research sheds new light on understanding glottal jet dynamics and the role of the jet on voice production. Specific questions being addressed center around whether and when the jet is quasi-steady. This is significant in the context of reduced order modeling of phonation; the governing equations can be simplified without loss of fidelity if the flow is quasi-steady [1-3]. This work builds on a framework [4-8] involving direct measurement of terms in the integral momentum (Newton's 2nd Law) and energy equations. Now, for the first time, simultaneous temporally and spatially resolved pressure and velocity measurements in a scaled up vocal fold model have allowed detailed examination glottal jets from the perspective of those equations. This enables exploration of issues such as how significant jet momentum and transverse forces on the glottal walls are to phonation.

Methods

Experiments were conducted using a 10x scaled-up model [4] in a free surface water tunnel. The combination of 10x physical scale, and low kinematic viscosity of water allows for a 1500x reduction in frequency to match Reynolds numbers and reduced frequencies of human phonation.

Two 2-D vocal fold models with semi-circular ends, 15 cm wide, 12.7 cm long, 27.3 cm high were placed 45.7 cm downstream of a 300 cm long, 27.3 cm square duct. Each model was computer driven at constant opening and clos-

ing speeds. The time from the start of opening to fully closed was defined as T_o. Once closed, the model was kept closed for an additional T_o for every cycle, so the full oscillation period was $2T_o$.

Digital Particle Image Velocimetry (DPIV), a 2-D flow measurement technique, coupled with time resolved pressure measurements along the vocal fold model, were made at four different Reynolds numbers from 3650 to 8100 and four oscillation frequencies from 0.035 Hz to 0.065 Hz. These correspond to human and reduced frequency ranges of ~50 Hz to ~110 Hz and 0.014 to 0.023, respectively.

Results and Discussion

The contribution of this work is demonstration of the significance of whether and when the glottal jet is quasi-steady. Earlier studies [5,6] of velocity-based quantities, *e.g.* maximum jet velocity and glottal volume flow rate, showed the jet is quasi-steady only for specific times in its evolution. Specifically, during jet formation and pinch-off, inertial effects are present. This is also seen in the time histories of transglottal pressure (TGP) shown in Figure 1; these are for the four different frequencies at Re = 7200. Note that while the model starts to open, is fully open, and closes at $t = 0$, $T_o/4$ and $T_o/2$, TGP minima and maxima do not occur at those times.

Figure 2 shows TGP *vs.* dynamic pressure including key time points as indicated in Figure 1. If the flow were quasi-steady, TGP should be equal to dynamic pressure; *i.e.* it

should follow the red line in Figure 2. Referring back to Figure 1, the time the jet is quasi-steady, between ② and ③, is only between $t/2T_o \approx 0.2$ and ~0.4 or ~40% of the vocal fold opening time. The time at which the jet pinches off, where sound production is most significant, is not quasi-steady.

This paper explores these issues in greater detail from the simultaneous velocity and pressure measurements and using related computational modeling work.

Acknowledgements

References

[1] Lighthill, M.J., Proc. Roy. A, 211:564-587, 1952.

[2] Titze, I.R. *Principles of Voice Production,* 2000

[3] Howe, M. & McGowan, R., Fluid Dynamics Res., 42:15001, 2010

[4] Krane, M.H. & Wei, T., JASA, 120:1578-1588, 2006.

[5] Krane, M.H, *et al*., JASA, 122:3659-3670, 2007.

[6] Krane, M.H., *et al*., JASA, 128:372-383, 2010.

[7] McPhail, M., *et al*., JASA, 146:1230-1238, 2019

[8] Sherman, E., *et al*., Fluid Dynamics Res., in print, 2019

corresponding author e-mail: twei3@unl.edu

Estimation of vocal fold physiology from voice acoustics using an artificial neural network

Zhaoyan Zhang*

Department of Head and Neck Surgery, UCLA School of Medicine, Los Angeles, CA, USA

Keywords: voice inversion; machine learning; voice simulation

Introduction

Many speech applications require estimating vocal fold properties from the produced acoustics. While there have been some previous research solving the inverse problem in voice production, they are often based on lumped-element models of phonation, whose model parameters are difficult to relate to realistic vocal fold properties. This study explores the feasibility of using machine learning methods to infer physiologically realistic vocal fold properties, including vocal fold length, thickness, depth, transverse and longitudinal vocal fold stiffness, vocal fold approximation, and the subglottal pressure, from the produced acoustics.

Methods

The data used in this study are from voice simulations in previous studies using a three-dimensional body-cover vocal fold model [1, 2]. It includes a total of 162,000 voice conditions, each simulates a half-second long voice production under different vocal fold geometric and mechanical properties and the subglottal pressure. For each condition, voice features are extracted from the glottal flow waveform and the output acoustics. Sixteen features are selected for this study, including fundamental frequency, sound pressure level, spectral measures of the voice spectra, and measures of the glottal flow waveform. The nine physiological measures to be estimated include the resting glottal angle, vocal fold vertical thickness, length, depths of the body and cover layers, transverse vocal fold stiffness, longitudinal stiffness of vocal fold cover and body, and subglottal pressure.

The data are randomly divided into three sets, each for training, validation, and testing, respectively. The training set was used to train a feedforward neural network, with different hidden layer configurations.

Results and Discussion

The results show that the subglottal pressure, vocal fold vertical thickness, and vocal fold length can be estimated with very low mean absolute errors. On the other hand, vocal fold stiffness, particularly stiffness in the body layer, consistently have large estimation errors. The performance for estimation of the initial glottal angle is somewhere in between.

This estimation performance is consistent with the findings in our previous parametric studies [1, 2]. These previous studies showed that the subglottal pressure, vocal fold vertical thickness, and length, which have better estimation performance in the present study, have consistent global effect on voice production. In contrast, the effect of vocal fold stiffness, particular the longitudinal stiffness, on voice production is less consistent, which may explain the poor estimation performance for these parameters in the present study.

Acknowledgements

References

[1] Zhang, Z., J Acoust Soc Am, 139: 1493-1507, 2016.
[2] Zhang, Z., J Acoust Soc Am, 142: 2311-2321, 2017.

corresponding author e-mail: zyzhang@ucla.edu

simVoice – Efficient acoustic propagation model of the human voice source using finite element method

Stefan Schoder[1]*, Alexander Hauser[1], Paul Maurerlehner[1], Sebastian Falk [2], Stefan Kniesburges[2] ,Michael Döllinger[2], Manfred Kaltenbacher[1]

[1]Institute of Mechanics and Mechatronics, TU Wien, Austria
[2]Department of Phoniatrics and Pediatric Audiology at the Department of Otorhinolaryngology Head & Neck Surgery, University Hospital Erlangen, Germany

Keywords: Computational aeroacoustics (CAA); finite element method (FEM); larynx; voice disorders

Introduction

The main objective of *simVoice* is the development of a three dimensional aero-acoustic numerical model for a prospective application in a clinical environment. The model consists of a computational fluid dynamics simulation (CFD) model with externally driven vocal folds motion, based on the Finite Volume method [1], and a computational aero-acoustic model (CAA), based on the 3D finite element method (FEM) [2] using the Perturbed Convective Wave Equation (PCWE). This contribution assesses the performance increase of a reference simulation model when changing discretization parameters while maintaining accuracy of the sound spectra in the acoustic far-field.

Methods

Based on the numerical model *simVoice*, the FEM submodel of the computational aeroacoustic simulation (CAA) is optimized, considering the computational efficiency. With respect to algorithmic efficiency at low Mach numbers, a hybrid aeroacoustic approach is applied, which performs in a first step an incompressible flow simulation on a computational grid which is capable of resolving all relevant turbulent scales. In a second step, the acoustic source terms on the flow grid are computed and a conservative interpolation to the acoustic grid is performed. In a third step, the PCWE is solved to obtain the acoustic field (see Figure 1).

Figure 1: Schematic of the flow diagram of a hybrid aeroacoustic workflow.

The CAA simulation setup consists of an unstructured meshed larynx and simplified vocal tract, a structured meshed propagation domain and perfectly matched layer (PML) region connected to the sub and supra glottal region (see Figure 2). Validation results from experiments are available for this setup. Starting from an initial optimal setup with respect to accuracy, the simulation model was tuned to reduce the computational effort.

Results and Discussion

First, the larynx and vocal tract regions mesh was coarsened. The reduction of the number of equations 195k to 46k in the two regions resulted in a reduction of the computational time from 3.6h to 3.1h. Second, in the PML region the reduction of the number of element layers has been investigated. Thereby total number of equations could be reduced from 243k to 134k by this measure and decreased the computational time from 3.1h to 46min. In a third step, the element size in the propagation region was increased too. The number of equations was lowered from 243k to 37k and the computational time diminishes to 8min.
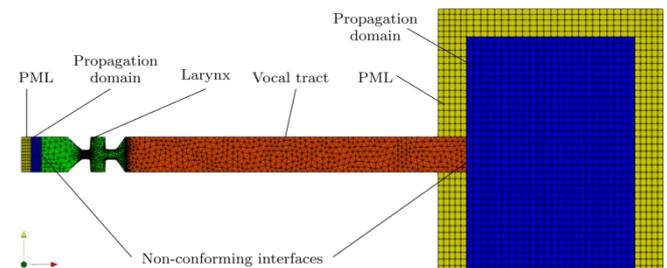
Figure 2: FEM model of the CAA simulation.

Conclusion

The demand for computational efficiency and fast computation is taken into account by the hybrid aeroacoustic approach in combination with the PCWE. Further optimization potential was exploited and the total computational time of the acoustic simulation is reduced to about 5% of the refined reference case.

Acknowledgments

References

[1] H. Sadeghi *et al*, J Voice, 2018.
[2] A. Hüppe *et al*,AIAA/CEAS , 2014.

*Stefan.schoder@tuwien.ac.at

Collagen fiber angles as a function of compression and depth within the nerve

Michael B Christensen[1,2]*, Reza Behkam[3], Melissa R Chao[4], Amanda C Stark[2], Jonathan P Vande Geest[3], Julie M Barkmeier-Kraemer[1]

[1]Department of Surgery, University of Utah, Salt Lake City, UT
[2]National Center for Voice and Speech, University of Utah, Salt Lake City, UT
[3]Department of Bioengineering, University of Pittsburg, Pittsburg, PA
[4]Department of Bioengineering, University of Utah, Salt Lake City, UT

Keywords: collagen; second harmonic imaging; compression

Introduction

The epineurium of peripheral nerves, of which collagen is a main constituent, serves as a protective layer for underlying nerve structures. It has previously been shown that collagen fibers within the epineurium align in the presence of uniaxial tension [1]. However, changes in collagen fiber angles as a function of compression have not previously been reported. Further, changes in collagen fiber angles as a function of depth, both in naïve and compressed nerves, has not been reported. Therefore, we investigated collagen fiber angles within the epineurial space both as a function of depth and compression in adolescent and juvenile pigs.

Methods

Recurrent laryngeal nerves (left and right) were harvested from both adolescent (3-4 months) and juvenile (1-3 weeks) pigs. Harvested nerves were cut into three adjacent sections, with two being compressed and the other serving as a non-tested control. Sections were compressed using a hydrostatic column in a custom built chamber. Following compression, nerves were fixed at pressure via fluid exchange using 10% formalin. Nerve diameters were measured using digital calipers. 100μm thick sections were collected using a cryostat and imaged using second harmonic generation (SHG) imaging with a Bruker Ultima multi-photon confocal microscope. Collagen fiber angles were evaluated using a custom Matlab code. Fiber angles were normalized so that the mean fiber angle for each section was equal to zero (zero being the direction along the length of the nerve).

Results

Non-tested sections from adolescent pigs had a mean diameter of 1.04mm, while compressed nerves had a mean diameter of 0.78mm (p=0.03). Non-tested sections from juvenile pigs had a mean diameter of 0.31mm, while compressed nerves had a mean diameter of 0.27mm (p=0.35).

Fiber angle measurements as a function of compression are shown in Figure 1A. Compression to the nerves from the adolescent pigs appeared to significantly alter the distribution of fiber angles, while compression to the nerves from juvenile pigs did not.

Examining collagen fiber angles as a function of depth, collagen fibers tend to be more aligned deeper in the nerve in adolescent pigs, both naively and after compression (Figure 1B), while depth did not appear to change fiber angle distributions in the juvenile pigs.
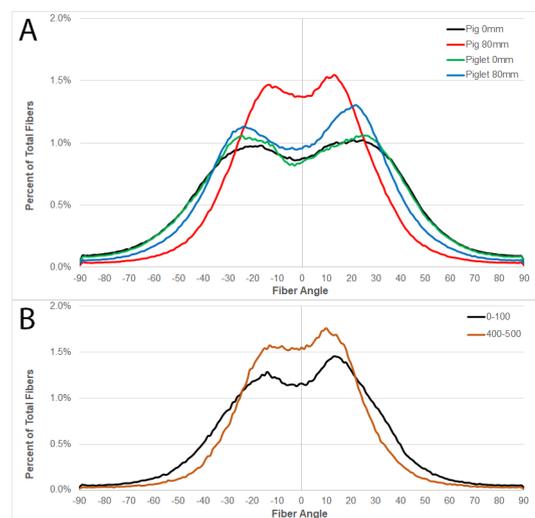

Figure 1: A) Compression seemed to significantly alter fiber angle distributions after compression in adolescent, but not juvenile, pigs. B) Deeper sections demonstrated fiber angle profiles that were more aligned than superficial sections both naively (not shown) and after compression (shown) in adolescent, but not juvenile, pigs.

Discussion

Compression significantly altered nerve diameter and collagen fiber angle distributions in adolescent, but not juvenile, pigs. This may be due to a more mature development of collagen in the older pigs, which allows the collagen to absorb compressive forces by aligning and becoming denser. The inability of collagen fibers to make similar changes in the young pigs suggests that these nerve may be more susceptible to compressive damage.

Funding

References

[1] Williams *et al*, J Biomech Eng, 136(8), 2014

*ben.christensen@utah.edu

The impact of phonomimetic vibration on vocal fold inflammation

David Hortobagyi*, Tanja Grossmann, Magdalena Maria Tschernitz, Magdalena Grill, Andrijana Kirsch, Claus Gerstenberger, Markus Gugatschka

Division of Phoniatrics, Department of Otorhinolaryngology, Medical University of Graz, Austria

Introduction:

To date there is no ultimate recommendation on how long patients should rest their voice after an acute injury. Orientating on other medical fields, where a rapid mobilisation after surgery is aimed at in order to achieve an earliest possible recovery. In the past years, there has been a tendency in decreasing the interval of voice rest after phono-microsurgery. The impact of mechanical forces on the lamina propria has been addressed by several papers, knowing that it plays an important role in phonation. [1][2][3][4][5][6]

This *in vitro* study aims to reveal the impact of mechanical stress by means of vibrational forces on vocal fold fibroblasts (VFF), the most abundant cells in the lamina propria, during an acute inflammatory reaction by putting an emphasis on the molecular and cellular level. [7]

Materials and Methods:

After an inflammatory stimulus, 144.000 human VFF per well were exposed to vibrational forces for three days in order to imitate a physiologic vocal strain. [8] Mechanical stimulation was applied for four hours over a period of 72 hours. An acute inflammatory reaction was simulated by adding 5ng/mL IL1b as well as the same concentration of TGFb1. Consequently, each experiment was carried out in four different settings: static as well as dynamic, both with and without cytokines.

Subsequently the differences in mRNA expression as well as the concentrations of proteins, being responsible for the extracellular matrix (ECM) composition of the lamina propria, and of proteins, which are known to play an important role in inflammatory reactions, were investigated and compared amongst the different groups. All experiments were repeated four times.

Results:

Cytokine treatment induced a change in the expression of ECM- as well as inflammation-related genes.

The pro-inflammatory cytokine IL11, as well as the myofibroblast marker alpha smooth muscle actin (α-SMA) were significantly reduced when additional vibration was applied. Hyaluronic acid (HA) concentration was significantly increased due to the cytokines, however mechanical stimuli did not show any effect on HA metabolism.

Discussion:

The upregulation of certain genes and proteins by the cytokine treatment was in line with other *in vivo* experiments. It therefore reflects the applicability of our in vitro inflammatory model. IL11 belongs to the IL6-family and appears to be an essential part in the down-streaming pathway of TGFb1. It therefore plays an important role in pro-fibortic reactions. [9]

During fibrogenesis VFF transform to myofibroblast, leading to a wound contraction in order to facilitate healing. However, simultaneously, they may impair oscillatory properties of the vocal folds and consequently cause hoarseness. [10] α-SMA is a well-established myofibroblast marker.[8][11] Elevated concentrations of these two proteins as well as the absence of changes in the desirable anti-fibrotic ECM component HA due to vibration, suggests a beneficial effect of mechanical stimulation following inflammation. These findings corroborate clinical studies which recommend early voice activation following an acute event. [5][12][13]

References:

[1] Guerra ML et al. Early mobilization of patients who have had a hip or knee joint replacement reduces length of stay in hospital: a systematic review. Clin Rehabil. 2015 Sep;29(9):844–54.

[2] Yakkanti RR et al. Impact of early mobilization on length of stay after primary total knee arthroplasty. Ann Transl Med. 2019 Feb;7(4):69–69.

[3] Kaneko M et al. Voice rest after laryngeal surgery: what's the evidence? Curr Opin Otolaryngol Head Neck Surg. 2017 Dec;25(6):459–63.

[4] Kaneko M et al. Optimal Duration for Voice Rest After Vocal Fold Surgery: Randomized Controlled Clinical Study. J Voice. 2017 Jan;31(1):97–103.

[5] Verdolini Abbott K et al. Vocal Exercise May Attenuate Acute Vocal Fold Inflammation. J Voice. 2012 Nov;26(6):814.e1-814.e13.

[6] Ishikawa K et al. Voice Rest Versus Exercise: A Review of the Literature. J Voice. 2010 Jul;24(4):379–87.

[7] Ling C et al. Alteration in cellular morphology, density and distribution in rat vocal fold mucosa following injury. Wound Repair Regen. 2010 Jan;18(1):89–97.

[8] Kirsch A et al. Development and validation of a novel phonomimetic bioreactor. PloS One. 2019;14(3):e0213788.

[9] Schafer S et al. IL-11 is a crucial determinant of cardiovascular fibrosis. Nature. 2017 Dec;552(7683):110–5.

[10] Branco A et al. Vocal fold myofibroblast profile of scarring: Scar Vocal Fold Myofibroblast. The Laryngoscope. 2016 Mar;126(3):E110–7.

[11] Jetté ME et al. Characterization of human vocal fold fibroblasts derived from chronic scar. The Laryngoscope. 2013 Mar;123(3):738–45.

[12] Whitling S et al. Absolute or relative voice rest after phonosurgery: a blind randomized prospective clinical trial. Logoped Phoniatr Vocol. 2018 Oct 2;43(4):143–54.

[13] Dhaliwal SS et al. Role of voice rest following laser resection of vocal fold lesions: A randomized controlled trial. The Laryngoscope. 2020 Jul;130(7):1750–5.

Inverse finite element of the aortic arch, implications for UVP patients

Gloriani Sánchez Marrero[1], Reza Behkam[1], Andrew J. Bierhals[2],
Julie M. Barkmeier-Kraemer[3], Jonathan Vande Geest[1]*

[1]Department of Bioengineering, University of Pittsburgh, Pittsburgh, PA, USA
[2]Institute of Radiology, Washington University School of Medicine, St. Louis, MO, USA
[3]Department of Surgery, University of Utah, Salt Lake City, UT, USA

Keywords: iFE; iUVP; aortic arch; RLN

Introduction

It is known that unilateral vocal fold paralysis (UVP) is a laryngeal disorder that occurs as a result of trauma, surgical intervention, respiratory infections or neurologic diseases. However, the etiology for several other cases remains unknown. The recurrent laryngeal nerve (RLN) is the motor supply of all the muscles of the vocal folds except for the cricothyroid. Damage to the RLN causes UVP. On the left side, the RLN loops around the aortic arch and other pulsating vessels which exposes the nerve to supraphysiological levels of stretch and compression. Previously, our research group found that the compliance of the aortic arch at the level of the brachiocephalic vein was higher in iUVP patients than in their age-gender matched controls [1]. Compliance of the aortic arch is greatly influenced by the biomechanical properties of the arch. In this study, we were interested in investigating the difference of these mechanical properties between iUVP patients and their matched controls.

Methods

We used an inverse finite element (FE) approach to model the patient-specific mechanical properties of the aortic arch in 10 iUVP patients and 10 matched controls. Gated MR images were taken of each subject over one cardiac cycle. With these, we defined geometrical and displacement boundary conditions. A digital volume correlation (DVC) algorithm was utilized to determine the displacement field that defined our boundary conditions. The aortic arch was described as an anisotropic hyperelastic tissue and the Holzapfel model was used to solve our simulation. Fiber orientation was defined from our prior work using small-angle light scattering [2]. The particle swarm optimization method was used to iteratively solve the FE model and calculate the unknown material parameters. The objective function was optimized to minimize the difference between the computed aortic arch diameter and the diameter measured using the gated MRI data at the level of the brachiocephalic vein.

Results

As shown in Figure 1 (top), matrix stiffness increased with age on both groups (iUVP and controls). In addition, for the old age group, iUVP patients had a smaller matrix stiffness than the healthy subjects. Similarly, in Figure 1 (bottom), fiber stiffness increases with age on the iUVP and control

groups. Also, collagen fiber stiffness is significantly smaller in old iUVP patients than in the healthy controls.
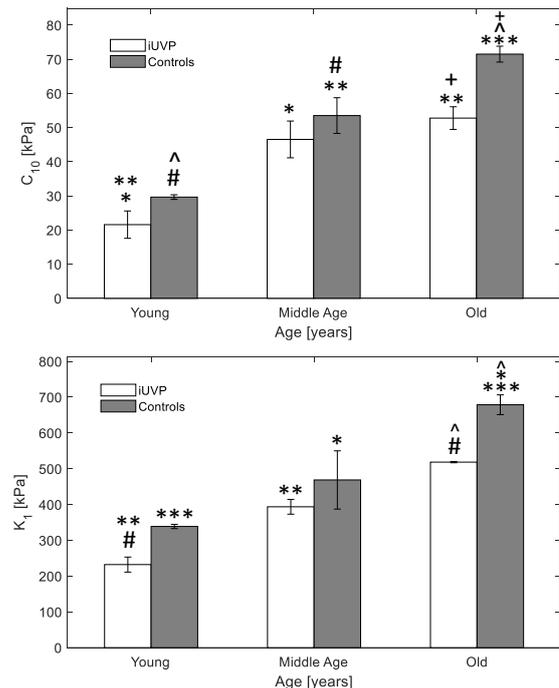
Figure 1: Matrix stiffness, C_{10} (top). Fiber stiffness, K_1 (bottom).

Discussion

Our results showed that matrix and fiber stiffness increased with age, which agrees with previous studies. More importantly, our results present smaller matrix and fiber stiffness in old iUVP patients compared to their age-gender matched controls. This suggests a lack of load-bearing, stiff structural components that might allow for supraphysiological loads on the RLN. These loads that generate tension or compression on the RLN might cause irreversible damage to the RLN.

Acknowledgements

References

[1] Behkam *et al*, JAP, 123:303-309, 2017.
[2] Haskett *et al*, BMM, 9:725-736, 2010.

The effect of electronic cigarette and tobacco smoke exposure on vocal fold mucosa remodeling and inflammation

Vlasta Lungova [1], Susan L. Thibeault [1]

[1]Department of Surgery, University of Wisconsin Madison,Wi, USA

Keywords: vocal fold mucosa; vaping; inflammation; remodeling

Introduction

Electronic cigarettes (e-cigarettes) are nicotine delivery systems that generate an inhalable aerosol containing nicotine, flavors, propylene glycol, and vegetable glycerin. They appeal to current and former smokers to help smoking cessation, and young people who never smoked. Emerging evidence suggests that there is a long-term risk associated with the use of e-cigarettes. At the moment it is difficult to quantify this risk due to the lack of studies, especially in relation to tobacco smoking. The main aim of this study was to investigate the effect of e-cigarettes on the inflammatory response and remodeling of human vocal fold (VF) mucosa.

Methods

Engineered human VF mucosae composed of human induced pluripotent stem cell derived VF epithelial cells and primary VF fibroblasts were exposed to 5% e-cigarette vapor extract (ECVE) for 1 week to mimic chronic e-cigarette vapor exposure. The effect of EVCE on the VF mucosa remodeling was examined by looking at the distribution and expression levels of structural and functional genes in VF epithelial cells and VF fibroblasts.

We also measured the levels of pro-inflammatory genes in these cell populations. Parallel experiments using 5% cigarette smoke extract were performed as controls.

Results

Human VF mucosa undergoes remodeling in response to 5% CSE and ECVE. As previously suggested, CSE causes excessive accumulation of cytokeratins in the epithelial luminal compartment, so called hyperkeratosis, in an attempt to protect epithelial cell layers from breakdown. On contrary, ECVE does not seem to induce keratotic changes in VF epithelial cells and reduces the cell adhesion, which may affect the compactness of the epithelial barrier. As a result, the epithelial cell layers may be more susceptible to damage. Moreover, oil droplets accumulate in the luminal cell layer, which likely induces inflammatory response in VF epithelial cells and fibroblasts.

Conclusion

ECVE compromises the compactness of the VF epithelial protective barrier. This raises concerns over the safety of e-cigarette use.

corresponding author e-mail

Pharyngoesophagial Segment Behavior in Esophageal and Tracheoesophageal Speech: A Review of Determining Aspects of Vocal Outcome

Ana Carolina Ghirardi[1]*, Andrey Ricardo da Silva[2]

[1]Department of Speech-Language Pathology and Audiology, Federal University of Santa Catarina, Florianópolis, SC, Brazil
[2]Department of Mechanical Engineering, Laboratory of Vibrations and Acoustics, Federal University of Santa Catarina, Florianópolis, SC, Brazil

Keywords: voice quality; laryngectomy; pseudoglottis; voice rehabilitation

Introduction

The Pharyngoesophagial Segment (PES) provides the voice source for totally laryngectomized patients. Variations in this structure's geometric and mechanical properties account for the difficulty in understanding its behavior during speech, crucial in providing clinical guidelines in voice rehabilitation, to predict therapeutic success and improve overall vocal outcome. As the fibers of the PES vibrate in both Esophageal (ES) and Tracheoesophageal (TES) speech, studies do not often compare this structure's behavior during the two forms of speech, considering them to occur in a similar manner. However, given that the *source of air* for voicing is different in ES e TES, it is reasonable to assume that the PES may behave differently in these two types of speech, accounting for the widely recognized difference in vocal outcomes and therapeutic success. Most studies focus on esophageal speech production and understanding the behavior of the PES during tracheoesophageal speech may aid in clinical management of speech rehabilitation of laryngectomized patients who use a tracheoesophageal prosthesis, and in understanding why the vocal failure rate is still fairly high in these cases. Thus, the **aim** of this paper is to conduct a discussion based on a literature review of the characteristics of the PES that directly influence speech in esophageal and tracheoesophageal speakers.

Methods

Literature Review Study, based on the guideline question: *How do the shape, position, pressure, tone, number of vibrating segments and symmetry of the mucosal wave of the PES during speech influence vocal outcome in esophageal and tracheoesophageal speakers?* A search for papers using predetermined keywords and boolean operators was conducted on the Scopus, Web of Science, Medline, and Pubmed databases. Papers were screened by title and abstract and those repeated in multiple databases, not available in either English, Portuguese, Spanish, French or Italian or did not address the guideline question were eliminated. Papers that met the predetermined criteria were read entirely and those that addressed the guideline question were included in the final sample for analysis. All steps were undertaken separately by two researchers and any doubts regarding the inclusion or exclusion of papers was solved by a third reader. There was no limitation regarding dates of publication.

Results and Discussion

The search yielded a total of 828 studies. 804 were excluded after all steps were completed. 24 studies were included for providing insight regarding the influence of specific characteristics of the behavior of the PES on voice production, assessed during speech. Different methods were used depending on the date of publication and aim, including x-ray, videofluoroscopy, esophageal insufflation, esophageal manometry, and stroboscopy. All studies related specific aspects of PES behavior to speech outcome, and mostly, for analysis, speakers were divided by an SLP into two separate groups: *good* and *poor* speakers, using different methods and scales, mostly based on perceptual criteria. There was little differentiation between overall speech performance and/or intelligibility and voice quality which, frequently, prevented more detailed analyses on voice. There may be a slight change in position of the PES (generally upward) in both ES and TES. In different voicing tasks (high/low pitch and strong/soft voice), there are changes in the shape of the PES, particularly in ES. Muscle tone plays an important role in voice onset in ES and in voice quality and phonation time in TES. Symmetric mucosal waves yield less hoarse and/or breathy voice quality for both groups. PES shape was not related to vocal outcome in ES, but, in TES, circular or oval PES tended to yield 'better' voice quality. Studies agree that pressure in the PES plays a major role in both ES and TES. In TES, pressure distribution along the esophagus is more significant in voice than pressure at rest at the upper esophageal sphincter, an important parameter for esophageal speakers. There is need for more specific criteria in assessing vocal quality in ES and TES so that surgeons, doctors and voice clinicians can better predict vocal outcomes, manage eventual setbacks and improve overall voice quality for laryngectomized patients.

Acknowledgements

References

[1] Búa *et al*, Annals of otology, Rhinology and Laryngology, 126(2):138-145, 2017.

[2] Isman *et al*, Head and Neck, 1992; 352-358.

[3] van As *et al*, Arch Otolaryngol Head Neck Surg 1999; 125: 891-897

[4] Aguiar-Ricz *et al*, J Voice 2007; 21(2): 248-256

*carolina.ghirardi@ufsc.br

*carolina.ghirardi@ufsc.br

Determination of an equivalent torsion spring constant for the valve flap of a tracheoesophageal voice prosthesis

Fernando H. T. Santos[1]*, Bárbara Rech[1], André M. C. Tourinho[1], Andrey R. da Silva[1]

[1]Department of Mechanical Engineering, Federal University of Santa Catarina, Florianópolis, SC, Brazil

Keywords: voice prosthesis; tracheoesophageal voice; voice aerodynamics

Introduction

Tracheoesophageal voice prostheses are the best voice recovery technique available for patients who have had their larynges removed [1]. In order to create accurate computer simulations of the fluid dynamics in the tracheoesophageal system, the voice prosthesis behavior needs to be properly modeled. By means of an experimental setup, the opening angle of a hinged flap commercial prosthesis was related to the trans-device pressure drop. The results were used as input parameters in simulations to obtain the force exerted by the airflow on the flap, so that an equivalent torsion spring constant could be determined for the hinged flap mechanism.

Methods

Using an experimental set-up (Figure 1), the prosthesis' flap opening angle was measured for each pressure drop.
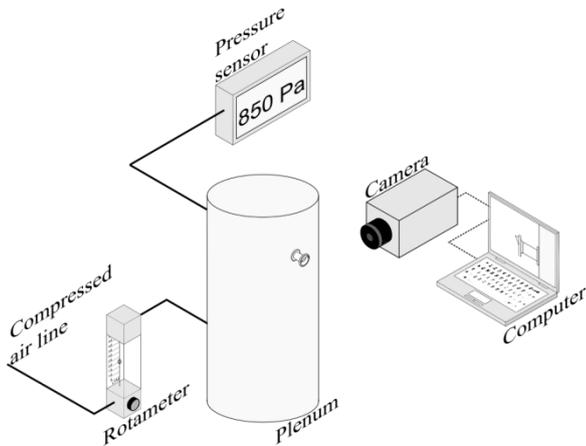
Figure 1: Experimental apparatus used to determine the opening angle of the prosthesis' valve mechanism.

Assuming that the prosthesis behavior is quasi-steady, the opening angles were used to produce geometries for models based on steady finite volume (FVM) and finite element (FEM) simulations, which were used to obtain the force exerted by the airflow on the prosthesis flap. The resulting forces were used to calculate the torque applied on the flap hinge.

Torsional springs within the elastic limit follow Hooke's law, which states that the angular displacement is directly proportional to applied torque. However, most off-the-shelf hinged voice prostheses are manufactured with a pre-tension to increase the resistance to low pressure airflows and avoid leakage [2]. Hooke's law is then written as

$$\tau = \kappa\theta + \tau_0, \qquad (1)$$

in which τ [Nm] is the torque, κ [Nm/rad] is the torsion spring constant, θ [rad] is the angular displacement, and τ_0 [Nm] is the minimum torque required to open the valve mechanism. The data set from the simulations was then processed using the least squares method to find the curve that best fits the results.

Results and Discussion

Figure 2 shows the results obtained and the adjusted curves. For the FVM data set the obtained torsion spring constant was found to be $\kappa = 4.23 \times 10^{-5}$ Nm/rad and the minimum opening torque $\tau_0 = 8.21 \times 10^{-6}$ Nm. And for the FEM data set the results were $\kappa = 3.73 \times 10^{-5}$ Nm/rad and $\tau_0 = 8.92 \times 10^{-6}$ Nm.
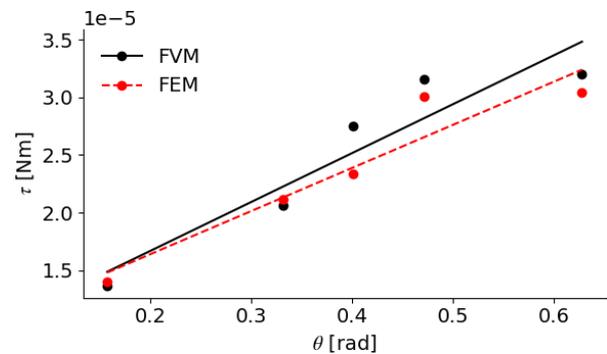
Figure 2: Torque by opening angle curves for FVM e FEM.

Both methods yield very similar results for the torsional spring constant and the minimum opening torque. These results can be used in more complex models to simulate the airflow in the tracheoesophageal system without relying on quasi-steady assumptions.

Acknowledgements

References

[1] Zenga *et al*, Oral Oncology, 86:38–47, 2018.
[2] Hilgers *et al*, The Laryngoscope, 100:1202–1207, 1990

The application of kinesiology tape as a tactile feedback for management of lower larynx position: a pilot study

Pedro Amarante Andrade[1]*, Marianne Bos-Clark [2], Janaina Pimenta[3], Sarah Catlow[4]

[1] Musical Acoustics Research Centre, Music and Dance Faculty, Academy of Performing Arts in Prague, Czechia
[2] Royal Devon and Exeter NHS Foundation Trust, Exeter, UK
[3] Espaço da Voz, Belo Horizonte, Brazil
[4] Plymouth Marjon University, Plymouth, UK

Keywords: Kinesio Taping; voice therpay; vertical larynx position

Introduction:

Kinesiology tape (KT) therapy has successfully been incorporated into speech pathology clinical practice for treatment of myofunctional orofacial disorders [1]. The elastic property of the tape allowed for support without hindering movement range [2]. Subjects with hyperfunctional voice disorders often present with a high held larynx position [3]. Therefore, the aim of this study is to assess the possibility of using an adapted application of KT as a tool for the management of the vertical larynx position (VLP).

Method:

Twenty-eight normophonic individuals were submitted to acoustic, auditory-perceptual and electroglottographic analyses (EGG) of voice pre and post KT application. Still images were taken to measure changes in the VLP. Sustained vowel [ɜ] and a singing passage ("Happy Birthday") were used as outcome measures. The KT application was done with the aim to promote a lower larynx position during phonation, mimicking the effects of the laryngeal depressor muscles.
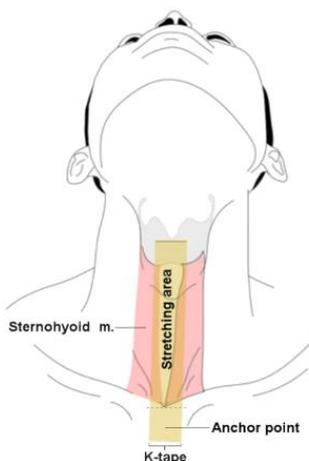
Figure 1: KT application.

Results:

LTAS analysis reviewed statistically significant changes for Total SPL (p < 0.01), Speaker's formant (p = 0.01), Alpha ratio (p = 0.04) and High/Low (p = 0.02) between conditions for the spoken task. No significance was found for the singing task, auditory-perceptual or CQ analyses. Due to the placement of the EGG electrodes the still images were not deemed reliable for assessing of the VLP. Formant analysis using FFT showed an overall drop in all formant frequencies apart from the fifth formant for the KT condition.

Discussion and Conclusion:

Based on the acoutic analysis, the results showed improved outcomes for the KT condition. Changes in VLP were not directly measured, however lower values for formant frequecies may be associated with a lower position of the larynx. Considering the small effect size expected for KT application with normophonics it seems reasonable to conclude that KT is a promising resource for aiding the management of the vertical larynx position. The use of KT as a therapeutic tool is not suggested as a replacement for speech therapy, but rather a possible resource to improve effectiveness of treatment.

Acknowledgements

References

[1] Bae Y. Change the Myofascial Pain and Range of Motion of the Temporomandibular Joint Following Kinesio Taping of Latent Myofascial Trigger Points in the Sternocleidomastoid Muscle. J Phys Ther Sci. 2014;26(9):1321-1324.

[2] Kenzo K, Hashimoto T, Tomoki O. Development of kinesio taping perfect manual. Kinesio Taping Assoc. 1996;6-10:117-118

[3] Angsuwarangsee T, Morrison M. Extrinsic laryngeal muscular tension in patients with voice disorders. *J Voice*. 2002;16(3):333-343. doi:10.1016/s0892-1997(02)00105-

* pedro.andrade@hamu.cz

Author Index

THANK YOU!

COMMITTEES

SCIENTIFIC COMMITTEE

Nathalie Henrich Bernardoni, PhD (Co-chair) – France
Lucie Bailly, PhD (Co-chair) – France
Dimitar Deliyski, PhD (ICVPB 2018) – USA
Matías Zañartu, PhD (ICVPB 2016) – Chile
Ingo Titze, PhD (ICVPB 2014, 2002, 1992, 1984) – USA
Zhaoyan Zhang, PhD (ICVPB 2014) – USA
Scott Thomson, PhD (ICVPB 2014) – USA
Susan Thibeault, PhD (ICVPB 2014, 2010) – USA
Michael Döllinger, PhD (ICVPB 2012) – Germany
Diane M. Bless, PhD (ICVPB 2010) – USA
Anne-Maria Laukkanen, PhD (ICVPB 2008) – Finland
Ken-Ichi Sakakibara, PhD (ICVPB 2006) – Japan
Antoine Giovanni, PhD (ICVPB 2004) – France

Philippe Dejonckere, PhD (ICVPB 2004) – The Netherlands
Hanspeter Herzel, PhD (ICVPB 1999) – Germany
Chuck Larson, PhD (ICVPB 1997) – USA
Ronald C. Scherer, PhD (ICVPB 1984) – USA
Pascal Perrier, PhD – France
Paul Luizard, PhD – Germany
Fabrice Silva, PhD – France
Thomas Hélie, PhD – France
Thierry Legou, PhD – France
Aude Lagier, PhD – Belgium
Sébastien Laporte, PhD – France
Emilie De Brosses, PhD – France

LOCAL ORGANIZING COMMITTEE

Nathalie Henrich Bernardoni, PhD (Co-Chair) - CNRS, GIPSA-lab
Lucie Bailly, PhD (Co-Chair) - CNRS, 3SR Lab
Laure Cavailles - Insight Outside
Paul Luizard, PhD - TU-Berlin
Fabrice Silva, PhD - CNRS, LMA
Thomas Hélie, PhD - CNRS, IRCAM
Thierry Legou, PhD - CNRS, LPL
Aude Lagier, PhD/ENT - CHU Liège, Belgium
Laurent Orgéas, PhD - CNRS, 3SR Lab

Sabine Rolland du Roscoat, PhD - Univ. Grenoble Alpes, 3SR Lab
Ihab Atallah, PhD - Univ. Grenoble Alpes, CHU Grenoble
Philippe Chaffanjon, PhD - Univ. Grenoble Alpes, CHU Grenoble, GIPSA-lab / LADAF
Frédéric Berthommier, PhD - CNRS, GIPSA-lab
Franck Quaine, PhD - Univ. Grenoble Alpes, GIPSA-lab
Akila Mokhtari - Univ. Grenoble Alpes, GIPSA-lab
Xavier Laval - Univ. Grenoble Alpes, GIPSA-lab
Fred Da Cunha - Imprimerie Grenoble INP

ICVPB INTERNATIONAL ADVISORY BOARD

Brad Story, PhD (Chair) - USA
Ingo Titze, PhD - USA
Johan Sundberg, PhD - Sweden
Nathalie Henrich Bernardoni, PhD - France
Christine Shadle, PhD - USA
Paavo Alku, PhD - Finland
Chuck Larson, PhD - USA
Jan Svec, PhD - Czech Republic

SPECIAL THANKS

We thank all authors, keynote speakers, pre-course speakers and reviewers for their contributions.

©Jérome NARCY