Department of Experimental Physics
Faculty of Science
Palacký University Olomouc, Czechia

Biomechanics and acoustics of voice production

Doctoral Dissertation, May 2023

Author: Hugo Lehoux, MSc

Study program: Biophysics

Supervisor: Doc. RNDr. Jan G. Švec, Ph.D. et Ph.D.

Lehoux, Hugo

Biomechanics and acoustics of voice production

[Biomechanika a akustika tvorby hlasu]

Doctoral Thesis, Palacký University in Olomouc, Czech Republic – with an abstract in Czech

STATEMENT OF ORIGINALITY

I declare that I am the only author of this thesis and all sources are properly cited

In Olomouc: Signature

Table of contents

ACKNOWLEDGMENTS

Personal thanks:

First and foremost, I would like to express my deepest gratitude to my supervisor Jan Švec, for putting his trust in me and for giving me the opportunity to work with him in the Voice Research Lab. During those five years of my doctoral studies, Jan Švec supported me in a very professional, yet friendly way, and guided me to grow as a person and as a researcher. I would also like to thank him for all the musical moments and discussions we shared.

During my Ph.D. studies, I had the opportunity to meet many other great researchers who also supported me in many ways. My most sincere thanks go particularly to Sten Ternström, who gave me the chance of going on a three-month research stay at the KTH to work with him, as well as Christian Herbst, Philipp Aichinger, and Lisa Popeil, for all their friendly support towards my work and career, for all the stimulating discussions about voice science, and for encouraging me to continue into the voice research field. I would also like to extend my thanks to Michael Döllinger, who allowed me to come to the University Clinic in Erlangen during two research stays to work on the analysis of hemilarynx experimental data.

I would not be where I am now without the immense help and support from my previous mentors. I will forever be grateful to Joe Wolfe, who welcomed me in his lab at the USNW in Sydney, during my 5-month internship as part of my master's degree, and to Joël Gilbert, who sadly passed away in 2022. Joël Gilbert never stopped believing in me and provided me with a unique chance to discover the wonderful world of music acoustics, and indirectly to voice research, when I was still a master student at Le Mans University. Let this thesis be a homage to his consistent passion and dedication.

Even though I spent part of my Ph.D. as the only student in the office, I appreciated the times when I was surrounded by my fellow colleagues of the Voice Research Lab. Many thanks to Ketaki Phadke, Petr Hájek, and Dominika Valášková for the friendly discussions and for making those times more fun.

As a musician and a singer, my time in Olomouc would not have been fulfilling without any musical encounters. I want to express my sincere thanks to the University choir Ateneo and its conductor Pavel Režný. The choir is a place where I grew as a singer and formed many friendships. Special thanks go to Luciáš for caring too much for others, and for encouraging me to be myself. I also feel very blessed to have met my bandmates Jeffrey and Kuba, who became close friends and with whom I had many musical experiences that I am proud of. I would like to express my deep thanks to them.

Finally, I would like to thank my family, all of my friends who supported me from France, and all the ones I met in Czech Republic.

Subglottal pressure oscillations in anechoic conditions:

I would like to acknowledge Dr. Ondřej Vencálek (Department of mathematical analysis and applications of mathematics, Faculty of Science, Palacký University in Olomouc) for his help with the statistical analysis, and my co-author Dr. Vít Hampala for his work on the experimental setup.

Investigation of frequency jumps in anechoic conditions:

I would like to recognize my co-author Dr. Martin Dobiáš for his help in acquiring the human excised larynges.

Analysis of chest and head registers:

I would like to thank the six listeners for their participation in the listening tests during this study.

Generation of synthetic high-speed videos:

I thank my co-author Dr. Pravin Kumar for his help with the publication of my work.

Support:

The work described in this thesis was supported by the following projects, grants and institutions:

- Palacký University students' projects IGA_PrF_2019_030 (1.3.2019-29.2.2020), IGA_PrF_2020_028 (1.3.2020-28.2.2021), IGA_PrF_2021_017 (1.3.2021-28.2.2022), and IGA_PrF_2022_029 (1.3.2022-28.2.2023)
- Czech Science Foundation (GA CR) project no. 19-04477S.

My three-month research stay at the Royal Institute of Technology (KTH), Stockholm, Sweden, was supported by the Erasmus Plus program and by the Palacký University.

In addition, my research stays in the Medical University in Vienna, Austria, and in the University Clinic in Erlangen, Germany, were supported by the OPIE project "Improvement of Doctoral Student Grant Competition Schemes and their Pilot Implementation", reg. no. CZ.02.2.69/0.0/0.0/19_073/0016713.

LIST OF PUBLICATIONS IN THIS THESIS

I. Lehoux, H., Hampala, V., & Švec, J. G. (2021). Subglottal pressure oscillations in anechoic and resonant conditions and their influence on excised larynx phonations. *Scientific Reports, 11*(1), 28. doi:10.1038/s41598-020-79265-3
Cited in: Švec et al. 2021; Lester-Smith et al. 2021; Herbst et al. 2023;

Authors' contribution: H.L.: formal analysis, investigation, data curation, writing—original draft, visualization. V.H.: methodology, subglottal tract model development, investigation. J.G.Š.: conceptualization, methodology, resources, writing—review and editing, supervision, funding acquisition.

Received **3rd place in the Student Scientific competition for the Dean's Award – 2020**) at Přírodovědecká fakulta UP, 17. listopadu 12, Olomouc, in the Physics section under doctoral division.

The work from this publication was presented by the author during the following events:

- Scientific Day "Acoustique de la voix et de la parole : modèles, mesures, maquettes [Acoustics of voice and speech: models, measurements, testbeds]" organized by the Speech Acoustics Group (GAP) of the French Acoustical Society (SFA) at the Institut de recherche et coordination acoustique/musique (IRCAM), Paris, France, 15/03/2023 - *Biomechanics and acoustics of vocal production: excised larynges, vocal registers and kinematic modelling* (Invited lecture, 30 minutes)
- Seminar at the Royal Institute of Technology (KTH), Stockholm, Sweden, 05/11/2021 - Ongoing and past activities at the Voice Research Lab, Palacký University, Olomouc, Czechia & Investigation of electroglottographic waveshapes using a kinematic model of the vocal fold vibrations (Lecture, 30 minutes)
- Czech Acoustical Society student seminar, 22/01/2021 (online) - Subglottal pressure oscillations in anechoic and resonant conditions and their influence on excised larynx phonations (Lecture, 15 minutes)
- e-Forum Acusticum, 7-11/12/2020 (online) - Development and Use of an Anechoic Subglottal Tract for Excised Larynx Experiments (poster presentation)
- 11th International workshop on Model and Analysis of Vocal Emissions for Biomedical Applications (MAVEBA), 17-19/12/2019 in Florence, Italy - Development and Use of an Anechoic Subglottal Tract for Excised Larynx Experiments. (Lecture, 15 minutes)

II. Lehoux, H., Herbst, C. T., Dobiáš, M., & Švec, J. G. (2023). Frequency jumps in excised larynges in anechoic conditions: A pilot study. *Journal of Sound and Vibration, 551,* 117607. doi: 10.1016/j.jsv.2023.117607

Authors' contribution: H.L.: Methodology, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization. CTH: Conceptualization, Investigation, Writing – review & editing. MD: Resources, Writing – review & editing. JGS: Conceptualization, Methodology, Investigation, Writing – review & editing, Supervision, Funding acquisition.

Received **3rd place in the Student Scientific competition for the Dean's Award – 2023**) at Přírodovědecká fakulta UP, 17. listopadu 12, Olomouc, in the Physics section under doctoral division.

The work from this publication was presented by the author during the following events:

- Scientific Day "Acoustique de la voix et de la parole : modèles, mesures, maquettes [Acoustics of voice and speech: models, measurements, testbeds]" organized by the Speech Acoustics Group (GAP) of the French Acoustical Society (SFA) at the Institut de recherche et coordination acoustique/musique (IRCAM), Paris, France, 15/03/2023 - *Biomechanics and acoustics of vocal production: excised larynges, vocal registers and kinematic modelling* (Invited lecture, 30 minutes)
- Czech Acoustical Society student seminar, Czech Technical University in Prague (ČVUT), Czechia, 19/01/2023 - Frequency jumps in excised larynges in anechoic conditions: A pilot study (Lecture, 15 minutes)

III. Lehoux, H., Popeil, L., & Švec, J. G. (2022). Laryngeal and Acoustic Analysis of Chest and Head Registers Extended Across a Three-Octave Range: A Case Study. *Journal of Voice (In press, early online)*. doi:10.1016/j.jvoice.2022.02.014
Cited in: Zita et al. 2022.

Authors' contribution: H.L.: formal analysis, investigation, data curation, writing —original draft, visualization. L.P.: conceptualization, data acquisition, writing—review and editing. J.G.Š.: conceptualization, methodology, data acquisition, resources, writing—review and editing, supervision, funding acquisition.

Received **2nd place in the Student Scientific competition for the Dean's Award – 2021**) at Přírodovědecká fakulta UP, 17. listopadu 12, Olomouc, in the Physics section under doctoral division.

The work from this publication was presented by the author during the following events:

- Scientific Day "Acoustique de la voix et de la parole : modèles, mesures, maquettes [Acoustics of voice and speech: models, measurements, testbeds]" organized by the Speech Acoustics Group (GAP) of the French Acoustical Society (SFA) at the Institut de recherche et coordination acoustique/musique (IRCAM), Paris, France, 15/03/2023 - *Biomechanics and acoustics of vocal production: excised larynges, vocal registers and kinematic modelling* (Invited lecture, 30 minutes)

- 7th International Physiology and Acoustics of Singing Conference PAS7+, 05/05/2022 (online) - Laryngeal and acoustic register differences across the entire pitch range of a female professional singer (Lecture, 15 minutes)
- Czech Acoustical Society student seminar, 20/01/2022 (online) - Laryngeal and acoustic analysis of chest and head registers extended across a three-octave range: a case study (Lecture, 15 minutes)
- Seminar at the Royal Institute of Technology (KTH), Stockholm, Sweden, 05/11/2021 - Ongoing and past activities at the Voice Research Lab, Palacký University, Olomouc, Czechia & Investigation of electroglottographic waveshapes using a kinematic model of the vocal fold vibrations (Lecture, 30 minutes)
- CHOICE FOR VOICE: Crossing boundaries in voice, 3-5/09/2021 (online) - Singing Register Differences in Vocal Fold Oscillations Observed Across Three Octaves Through Laryngeal High-Speed Videoendoscopy (Poster presentation)
- Viennese Workshop on Voice Quality, 27/08/2021 (online) - Chest versus Head Register Differences in Vocal Fold Oscillations across the Pitch Range of Three Octaves (Lecture, 15 minutes)
- 12th International Conference on Voice Physiology and Biomechanics (ICVPB), 2-4/12/2020 (online) - Singing Register Differences in Vocal Fold Oscillations Observed Across Three Octaves Through Laryngeal High-Speed Videoendoscopy (Poster presentation)

IV. Aichinger, P., Kumar, S. P., Lehoux, H., & Švec, J. G. (2022). Simulated Laryngeal High-Speed Videos for the Study of Normal and Dysphonic Vocal Fold Vibration. *Journal of Speech Language and Hearing Research, 65*(7), 2431-2445. doi:10.1044/2022_jslhr-21-00673

Authors' contribution: P.A.: conceptualization, software development, data acquisition, formal analysis, investigation, writing —original draft, visualization, funding acquisition. S.P.K.: software development, writing—review and editing. **H.L. software development, writing—review and editing**. J.G.Š.: conceptualization, writing—review and editing, supervision, funding acquisition.

ABSTRACT

Voice is a complex phenomenon, originating from the airflow modulated by the vibrations of the vocal folds, located in the larynx. The vibratory pattern of the vocal folds is influenced by the action of the intrinsic and extrinsic laryngeal muscles, which impacts the sound quality of the voice. The vocal folds also vibrate in a system comprising supraglottal and subglottal cavities presenting acoustic resonances, which can interact with the vocal fold vibrations. To better understand the mechanisms of voice production, the investigation of the vocal fold vibratory properties, and how they interact with the supraglottal and subglottal resonances, is therefore essential. This dissertation carries out this task by focusing on specific phenomena that are not sufficiently explained, such as the influence of the subglottal resonances on subglottal pressures and on the vocal fold vibrations, the occurrences of frequency jumps, and the notion of voice registers. This dissertation provides novel results, which help to improve our understanding of those phenomena, and can be divided into four main parts, related to four original articles.

In the first part, the development of an anechoic subglottal tract, to allow the observation of the vocal fold vibrations in subglottally anechoic conditions, is described. The acoustic response of the anechoic subglottal tract is first measured and compared with the acoustic response of a resonant subglottal tract, which reveals that the anechoic subglottal tract successfully suppresses its acoustic resonances. It is then used during excised deer larynx experiments without a supraglottal tract, where the subglottal pressure and the frequency of the vocal fold oscillation are measured, and compared with resonant conditions. In anechoic conditions, the subglottal pressure waveform resembles the theoretical source flow waveform, exhibiting a constant value during the closed phase, whereas it contains strong fluctuations during the closed phase in resonant conditions. Additionally, in resonant conditions the oscillation frequency of the vocal folds shows consistently lower values than in anechoic conditions. This indicates that the presence of subglottal resonances alters not only the subglottal pressures but also the vocal fold vibrations.

In the second part, the developed anechoic subglottal tract is further used to investigate frequency jumps in both anechoic and subglottally resonant conditions, using human excised larynges without a supraglottal tract. In both these conditions, larynges exhibit consistent frequency jumps, which suggests that supraglottal or subglottal resonances are not needed for the jumps to occur. This indicates that such jumps are primarily caused by inherent nonlinear-dynamic properties of the larynges. Nevertheless, some differences are observed in the jumps between anechoic and resonant conditions, indicating that the subglottal resonances also have an influence, although secondary.

In the third part, chest and head registers are investigated in vivo in a professional female singer using laryngeal high-speed videoendoscopy, electroglottography, and an acoustic analysis of a microphone signal of the radiated sound. Chest and head registers are commonly assumed to be limited to low frequencies and high frequencies, respectively, and are mostly studied within a limited frequency range where the production of both registers is comfortable. Here, the singer produced both chest and head registers across a wide frequency range. The aim of this study is to find relevant parameters of vocal fold vibrations that allow to discriminate between the two registers, regardless of frequency. The microphone sound is used in blind listening tests to determine whether the intended registers can be audibly distinguished and reveals that the register changes are not always perceivable by listeners. Kinematically, the singer consistently differed the chest register from the head register by means of a more rapid glottal

closure. This was usually, but not necessarily, accompanied by a greater membranous adduction of the vocal folds.

Finally, a kinematic model of the vocal fold vibrations for generating synthetic kymograms resembling those observed in vivo is extended to three dimensions, to allow the generation of synthetic high-speed videos of the entire glottis. This aims at helping to improve our interpretation of laryngeal high-speed videoendoscopy, especially in a clinical context where this technique is not yet frequently used. Synthetic high-speed videos corresponding to different types of phonation are generated. Specifically, the model is capable of generating videos exhibiting both regular vibrations, which are typical for healthy voices, and irregular vibrations, which are usually associated with voice pathologies. The extended model can be used to study the influence of the geometric and kinematic VF parameter variation on the appearance of the vibratory patterns, as observed through high-speed videos, kymograms, and phonovibrograms.

ABSTRAKT

Hlas je složitý jev, který vzniká prouděním vzduchu modulovaným vibracemi hlasivek umístěných v hrtanu. Vibrace hlasivek jsou ovlivňovány činností vnitřních a vnějších hrtanových svalů, což má vliv na kvalitu hlasu. Navíc, hlasivky kmitají v traktu zahrnujícím supraglotické a subglotické dutiny, které vykazují akustické rezonance, jež mohou interagovat s vibracemi hlasivek. Pro lepší pochopení mechanismů tvorby hlasu je proto nezbytné zkoumat vibrační vlastnosti hlasivek a jejich interakci se supraglotickými a subglotickými rezonancemi. Tato disertační práce plní tento úkol tím, že se zaměřuje na specifické jevy, které nejsou dostatečně vysvětleny, jako je vliv subglotických rezonancí na kmity hlasivek, na přeskoky frekvence hlasu a na hlasové rejstříky. Tato disertační práce přináší nové výsledky, které přispívají k lepšímu pochopení těchto jevů, a lze ji rozdělit do čtyř hlavních částí, které souvisejí se čtyřmi původními články.

V první části je popsán vývoj bezdozvukového subglotického traktu, který odstraňuje akustické rezonance a umožňuje pozorování vibrací hlasivek v bezdozvukových podmínkách. Akustická odezva bezdozvukového subglotického traktu je nejprve změřena a porovnána s akustickou odezvou rezonančního subglotického traktu. Výsledky ukazují, že vyvinutý trakt úspěšně potlačuje své akustické rezonance. Poté je tento trakt použit při experimentech s preparáty hrtanu, kde jsou subglotický tlak a frekvence kmitání hlasivek porovnávány mezi bezdozvukovými a rezonančními podmínkami. V bezdozvukových podmínkách se průběh subglotického tlaku podobá teoretickému průběhu zdroje hlasu a vykazuje konstantní hodnotu během fáze uzavření glotis, zatímco v rezonančních podmínkách obsahuje silné fluktuace během fáze uzavření glotis. V rezonančních podmínkách navíc frekvence kmitání hlasivek vykazuje konzistentně nižší hodnoty než v bezdozvukových podmínkách, což ukazuje, že přítomnost subglotických rezonancí mění nejen subglotický tlak, ale i kmitání hlasivek.

Ve druhé části je vyvinutý bezdozvukový subglotický trakt použit ke zkoumání přeskoků frekvence hlasu jak v bezdozvukových, tak v subgloticky rezonančních podmínkách, a to za použití lidských preparátů hrtanu, bez připojeného supraglotického traktu. V obou podmínkách vykazují hrtany konzistentně přeskoky frekvence kmitání hlasivek, což dokazuje, že supraglotické nebo subglotické rezonance nejsou pro vznik přeskoků nutné. To naznačuje, že tyto přeskoky frekvenci jsou primárně způsobeny inherentními nelineárně-dynamickými vlastnostmi hrtanu. Nicméně byly pozorovány i drobné rozdíly v přeskocích frekvencí mezi bezdozvukovými a rezonančními podmínkami, což naznačuje, že subglotické rezonance zde mají také vliv, i když druhotný.

Ve třetí části je studován hrudní a hlavový rejstřík u profesionální zpěvačky pomocí vysokorychlostní videoendoskopie hrtanu, elektroglotografie a akustické analýzy signálu z mikrofonu registrujícího vyzařovaný zvuk. Běžně se předpokládá, že hrudní a hlavový rejstřík je omezen na nízké, respektive vysoké frekvence, a tyto rejstříky se studují v omezeném frekvenčním rozsahu, kde lze pohodlně produkovat oba rejstříky. V našem případě zpěvačka produkovala jak hrudní, tak hlavový rejstřík v širokém frekvenčním rozsahu. Cílem této studie bylo nalézt relevantní parametry kmitání hlasivek, které umožňují rozlišit oba rejstříky bez ohledu na frekvenci. Zvuk z mikrofonu byl použit při zaslepených poslechových testech, pro zjištění, zda lze oba rejstříky poslechově rozlišit. Testy odhalily, že změny rejstříků nejsou posluchači vždy dobře rozpoznatelné. Kinematicky se hrudní rejstřík odlišoval od hlavového rychlejším uzavíráním glotis. To bylo obvykle, ale ne nutně, doprovázeno výraznější addukcí blanité části glottis.

V poslední části je vylepšen kinematický dvojrozměrný model kmitání hlasivek pro generování syntetických kymogramů, které se podobají kymogramům pozorovaným in vivo. Model je rozšířen do tří rozměrů, aby bylo možné generovat syntetické vysokorychlostní videozáznamy celé glottis. Cílem je pomoci zlepšit interpretaci záznamů z vysokorychlostní videoendoskopie hrtanu, zejména v klinickém kontextu, kde je tato technika využívána pouze zřídka. Pomocí modelu jsou generována syntetická vysokorychlostní videa odpovídající různým typům fonace. Je ukázáno, že model je schopen generovat videa vykazující jak pravidelné kmity, které jsou typické pro zdravé hlasy, tak nepravidelné kmity, které jsou obvykle spojeny s patologiemi hlasu. Rozšířený model lze použít ke studiu vlivu geometrických a kinematických změn parametrů VF na průběh kmitů, tak jak jsou pozorovány prostřednictvím vysokorychlostních videí, kymogramů a fonovibrogramů.

1. MOTIVATIONS FOR THE WORK OF THE AUTHOR

As a musician and a singer, I was always fascinated by the mechanisms of sound production in musical instruments, including the human voice. During my master's degree in Acoustics at Le Mans University, France, I had the chance of being introduced to the world of music acoustics through the realization of a research project, which further developed my desire to explore it. My amateur singing practice was also filled with questions and an inability to grasp the various abstract notions my teacher tried to explain to me: what actually is a register? Why does my voice break when I try to sing high notes in my chest voice? What, in my throat, is responsible for creating this or that sound? Consequently, I took the opportunity to satisfy my desire of knowledge and went on a 5-month research internship at the University of New South Wales, Sydney, Australia, to work with Prof. Joe Wolfe on the study of voice instabilities caused by acoustic interactions with the supraglottal tract. This internship allowed me to learn about the basic mechanisms of voice production, and after my master's degree I wanted to develop my knowledge on more fundamental aspects of voice production. While in Sydney, I read about the research work of Doc. Jan Švec on the investigation of voice breaks through excised larynx experiments, which encouraged me to come to the Palacký University in Olomouc, Czechia, to work under his supervision for my doctoral studies. My primary theme of research is the investigation of the influence of the subglottal resonances on the properties of the vocal folds, including voice breaks. During my course of study, I also had the opportunity to meet and work with the American singer Lisa Popeil on the topic of voice registers, which resulted in another scientific paper that I authored. Additionally, I came across the mucosal wave model developed by Doc. Jan Švec and Dr. Pravin Kumar at the Palacký University and I was fascinated by its capabilities. My encounter with Dr. Philipp Aichinger at the Medical University of Vienna, Austria, stimulated the desire to improve this model, which led to an additional scientific paper that I co-authored.

The present thesis can be therefore divided in four main parts. The first part focuses on the continuation of the development of an anechoic subglottal tract for excised larynx experiments, which is used to investigate the inherent vibratory properties of the vocal folds in anechoic conditions. The second part describes the use of the anechoic subglottal tract to investigate the occurrence of frequency jumps. The third part is devoted to the laryngeal analysis of chest and head registers through laryngeal high-speed videoendoscopy. Finally, the fourth part deals with the development of the mucosal wave model, particularly its extension to three dimensions, to allow the simulation of high-speed videos of the vibrating glottis. The following introduction provides a brief, non-exhaustive description of the laryngeal, subglottal, and supraglottal anatomy, as well as of the mechanisms of vocal fold oscillations, which are relevant for this thesis. Some of the techniques used to investigate voice production, in vivo, ex vivo, and through numerical modelling, are also reviewed.

2. INTRODUCTION

2.1. Fundamentals/basics of voice production

Voice production is a complex phenomenon which normally involves many organs in several parts of the body. Breathing muscles (intercostal muscles, abdominal muscles, diaphragm, etc.) create an overpressure in the lungs below the vocal folds (VFs – see Figure 1), which generates the power needed for voice production. This power is transmitted to the VFs in the larynx and causes them to oscillate, which is a process called phonation. The sound produced by the VF vibrations propagates below (in the trachea) and above (in the pharynx, oral cavity, and nasal cavity) the VFs, and is radiated through the lips and sometimes through the nostrils, when the velum opens (see Figure 1). The oral and nasal cavities act as filters on the sound generated by the VF vibrations, which modifies its timbre in order to create different vowels, for example. Speech commonly uses sounds which are either voiced (when the VFs vibrate) or unvoiced (where the VFs do not vibrate), however this thesis only focuses on the vibration of the VFs, and therefore will not discuss unvoiced sounds.

The space between the VFs is referred to as the glottis (see Figure 2), and therefore the cavities below and above the glottis are referred to as the subglottal (below the glottis) tract and the supraglottal (above the glottis) tract (also sometimes called the vocal tract). Figure 1 shows a schematic diagram of various organs used in voice production.

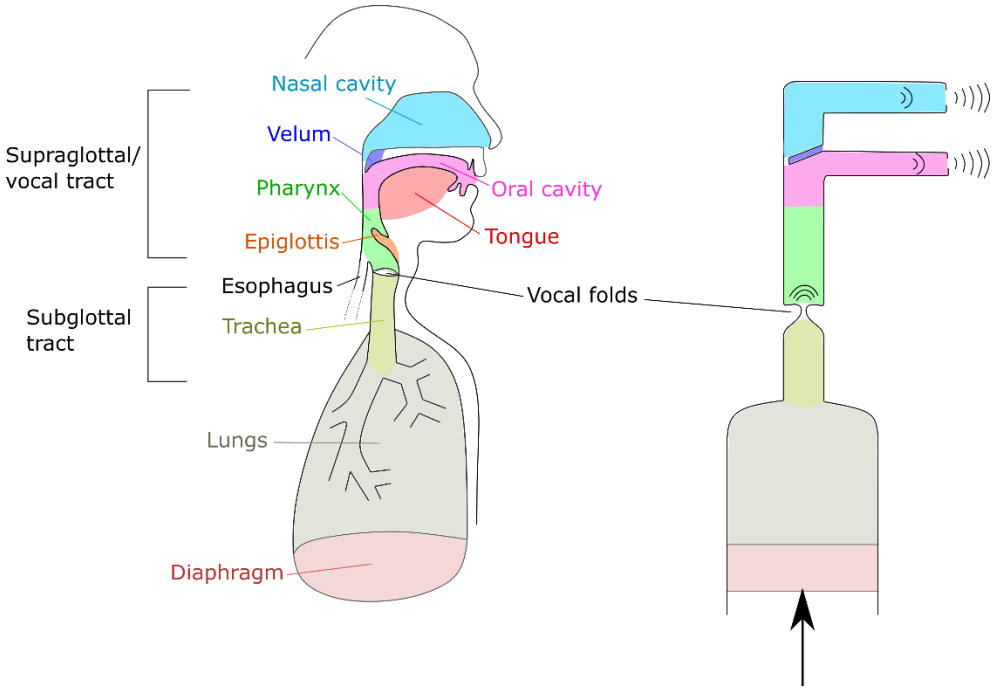


Figure 1. Basic anatomy of the phonatory system. Modified from Flanagan (1972).

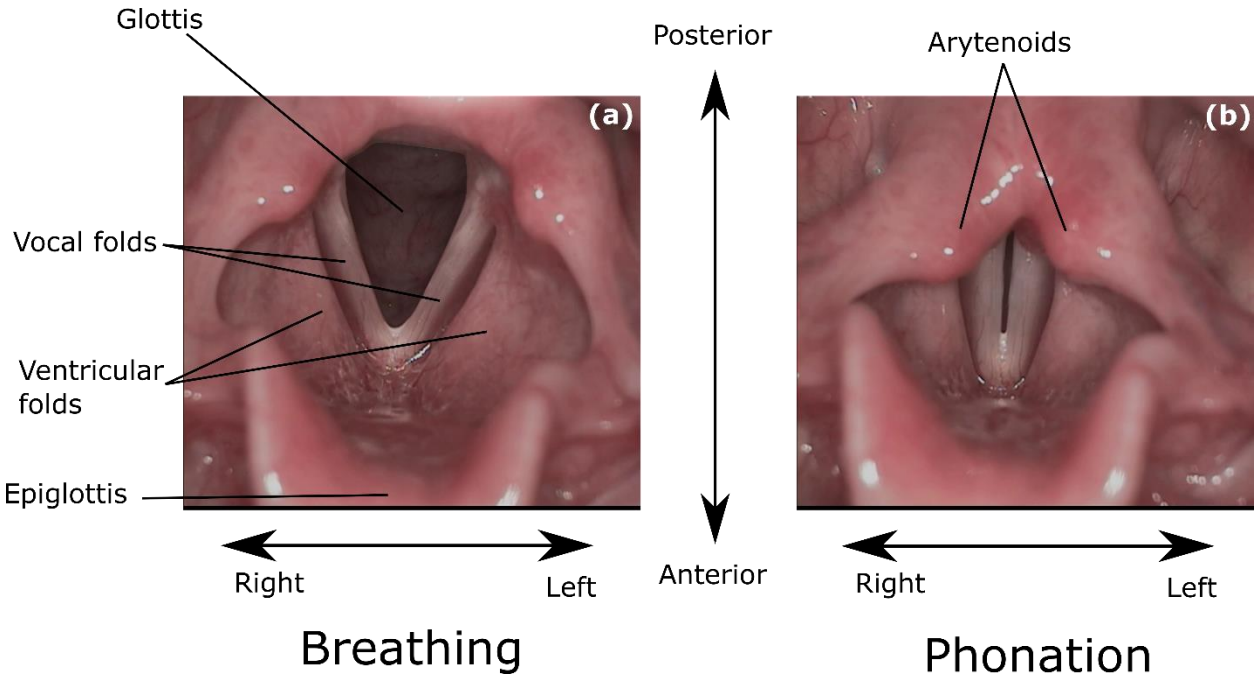


Figure 2. Laryngoscopic images showing the anatomy of the larynx. (a) During breathing; (b) During phonation. Images taken from laryngeal videos, courtesy of MUDr. Jitka Vydrová, Voice Centre Prague, Medical Healthcom Ltd, Czechia.

2.1.1. Anatomy of the vocal apparatus

2.1.1.1. Larynx

The VFs are attached anteriorly to the thyroid cartilage, and posteriorly to the arytenoids (see Figure 3a). The arytenoids are attached to the superior and posterior ridge of the cricoid cartilage (see Figure 3a). The thyroid cartilage is also attached to the cricoid cartilage through the cricothyroid articular joint (see Figure 3b), which allows the thyroid cartilage to rotate anteriorly and elongate the VFs through the action of the cricothyroid (CT) muscle (see Figure 3). The cricothyroid joint is not strictly a rotational axis, as there can also be some small translational (anterior-posterior) component to the movement, however this translational component has been shown to have a smaller influence than the rotational component (Vilkman, 1987), therefore we do not consider the translational component in this thesis.

The intrinsic laryngeal muscles described in Figure 4 are used to move and rotate the arytenoids with respect to the cricoid cartilage. The action of those muscles either leads to the abduction (parting away from the midline) or the adduction (approximation towards the midline) of the VFs. The abductor muscle is the posterior cricoarytenoid (PCA) muscle (Figure 4b), and the adductor muscles are the lateral cricoarytenoid (LCA) muscle (Figure 4c) and the interarytenoid (IA) muscle (Figure 4d) (Van den Berg et al., 1960; Titze, 2000a).

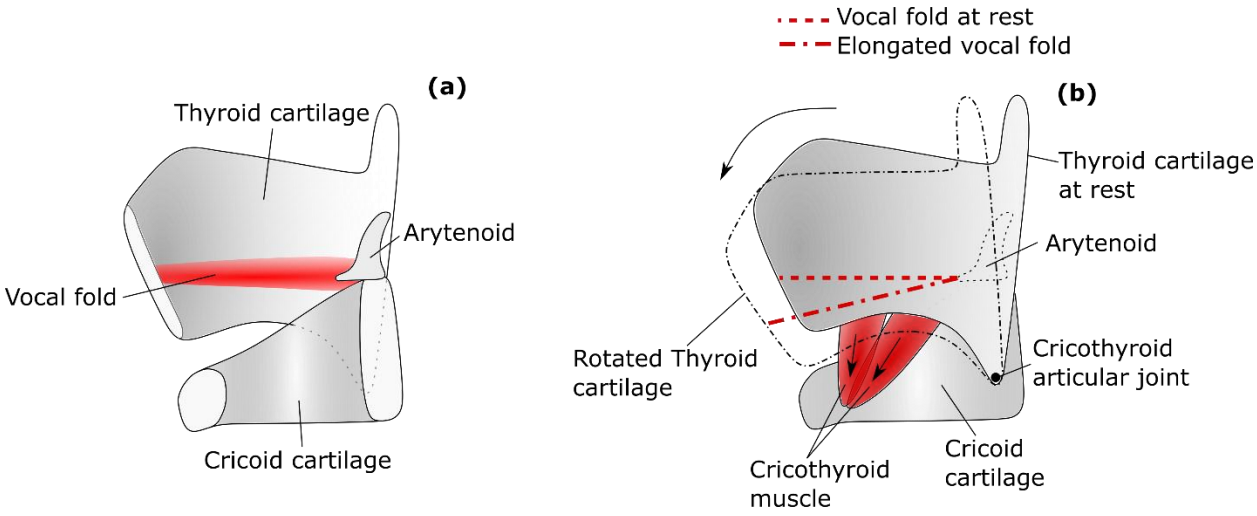


Figure 3. (a) Description of the laryngeal cartilages and the vocal fold attachment; (b) Description of the thyroid cartilage rotation and the vocal fold elongation, as a result of the CT muscle action. Modified from Titze et al. (1988), Fig. 1.

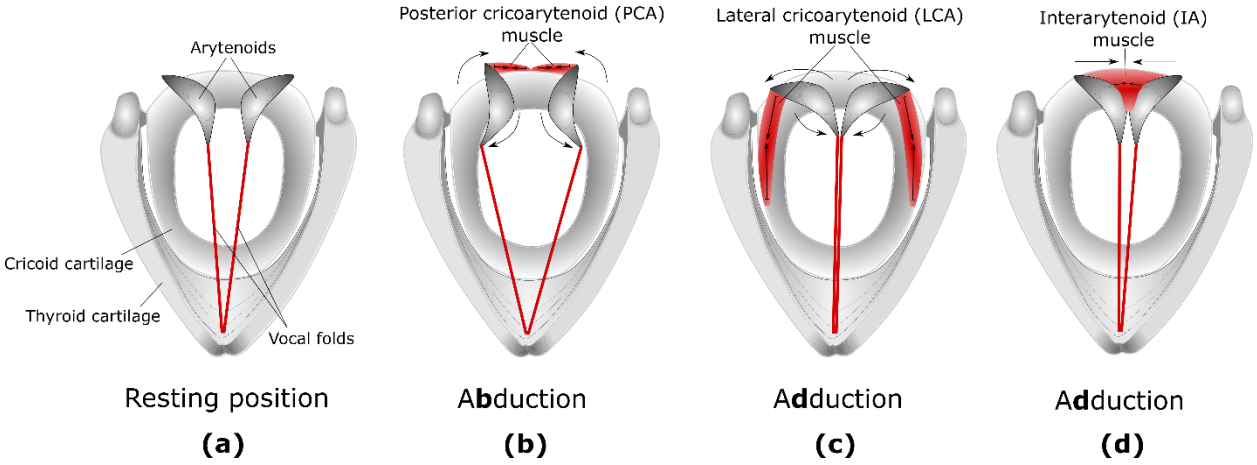


Figure 4. Description of the intrinsic laryngeal muscles used for abduction and adduction. Modified from Zemlin (1981).

VFs are complex tissues composed of five different layers (Hirano, 1974, 1975; Titze, 2000a): the epithelium, the superficial, intermediate, and deep layer of the lamina propria, and the thyroarytenoid (TA) muscle layer (see Figure 5). The layers have different compositions and therefore different biomechanical properties. They are sometimes grouped together into two main parts called the “body” (comprising the muscle layer and the deep layer of the lamina propria) and the “cover” (comprising the superficial and intermediate layers of the lamina propria, and the epithelium) (Hirano, 1974, 1975; Titze et al., 1988; Titze, 2000a).

Because the VFs are mainly composed of the TA muscle, their shape and biomechanical properties vary depending on the activity of this muscle. The contraction of the TA muscle will tend to actively shorten, tense, and bulge the VFs (see Figure 6) (Hirano, 1974, 1975; Wu et al., 2019). Alternatively, the action of the CT and LCA muscles will also passively tense the vocal fold tissues from their elongation.

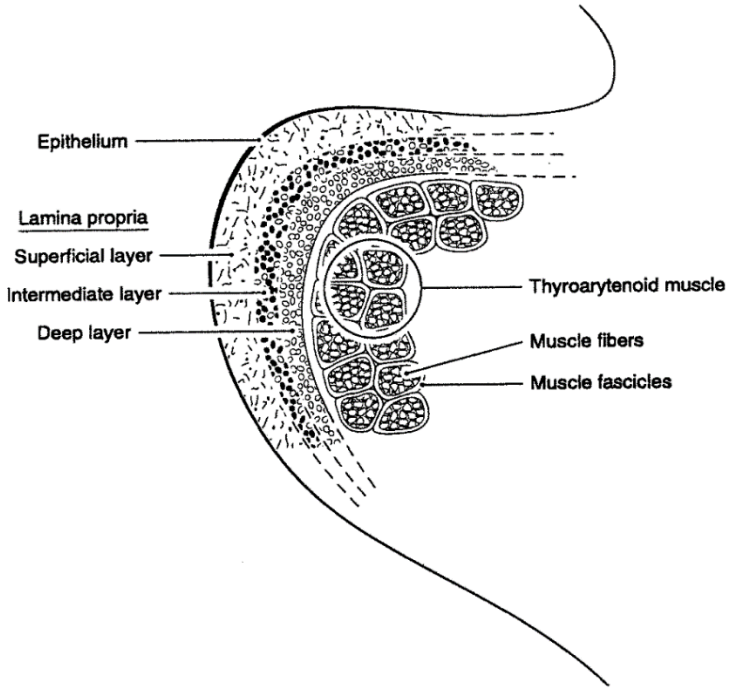


Figure 5. Coronal view of the VF and its different layers. Taken from Titze (2000a).

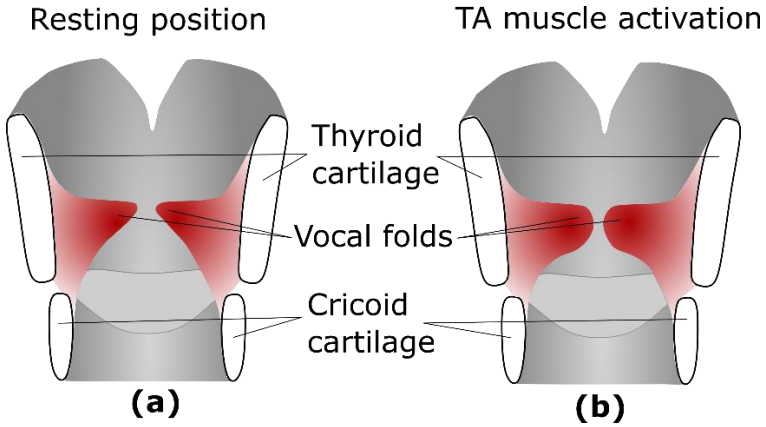


Figure 6. Effect of the TA muscle activation. (a) Resting position, the VFs are thin; (b) when the TA muscle activates, the VFs are bulged and thicker. Based on histological data from Hirano (1975).

2.1.1.2. Supraglottal cavities

The supraglottal tract (also called vocal tract) has an average effective length of about 17 cm (Baer et al., 1991; Story et al., 1996; Fitch et al., 1999; Vorperian et al., 2009), but its effective length can be increased or decreased by moving the larynx up (for example when swallowing) and down (for example when yawning). The supraglottal tract can be assimilated to a curved tube, through which the sound generated by the VF vibrations propagates and reflects at the mouth opening (as indicated on Figure 1). The sound can also propagate through the nasal cavity and reflect at the nostrils, but often the entrance to the nasal cavity is closed by the velum and the nasal cavity can then be disregarded. Constrictions in the supraglottal tract, caused by e.g. the tongue placement, can also cause the sound to reflect. Therefore, the acoustic transfer function of the supraglottal tract exhibits local maxima, which are associated with acoustic

resonances. The frequency of those resonances depend on the geometry of the supraglottal tract, particularly on its effective length and the position of its constrictions. Figure 7 depicts the supraglottal tract geometry and the corresponding acoustic transfer function, for two different English vowels.

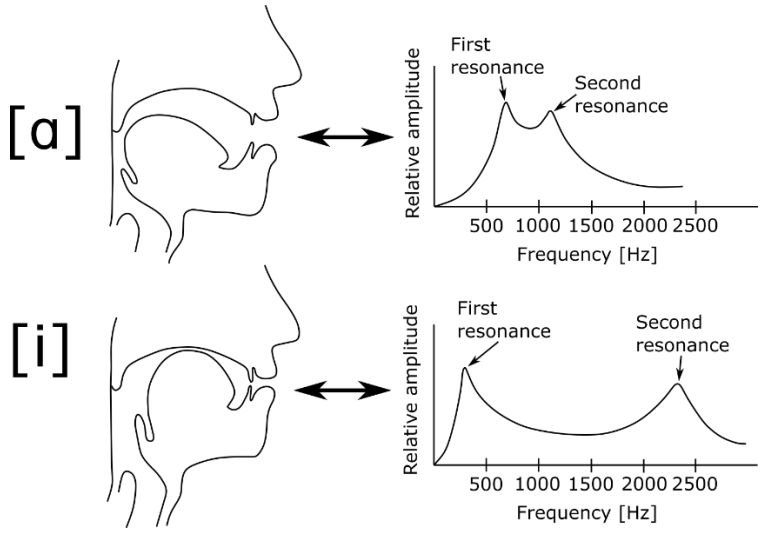


Figure 7. Description of the changes in the acoustic properties of the supraglottal tract caused by changes in its geometry. The two examples show the supraglottal tract geometry and the corresponding transfer function for the [a] and the [i] vowels. Modified from Titze (2000c).

2.1.1.3. Subglottal cavities

The subglottal tract consists of the cavity below the glottis, down to the point where the trachea splits into the bronchi. The length of the trachea, as measured from the lower border of the cricoid cartilage down to the inferior border of the carina (ridge of cartilage at the base of the trachea, separating the openings of the left and right main bronchi – see Figure 8) is, on average, about 12 cm (Grillo et al., 1964). This length can increase by up to 2 cm with deep inspiration (Holbert et al., 1995).

The trachea is rather straight and presents acoustic resonances f_{Rn} which approximately correspond to those of a straight tube open at one end and closed at the other end. Frequencies of such acoustic resonances can be expressed by the equation $f_{Rn} \approx (2n - 1)f_{R1}$, where n is an odd integer. Cranen et al. (1987) measured average values of f_{R1} = 510 Hz and f_{R2} = 1355 Hz, while other studies found comparable values (Fant et al., 1972; Lulich et al., 2012; Sundberg et al., 2013). The soft walls of the trachea create a large acoustic damping due to viscous losses, which is reflected in the high bandwidth (>100 Hz, Q factor > 0.2 for the first resonance) of the subglottal resonances, as measured by e.g., Cranen et al. (1987) and Fant et al. (1972).

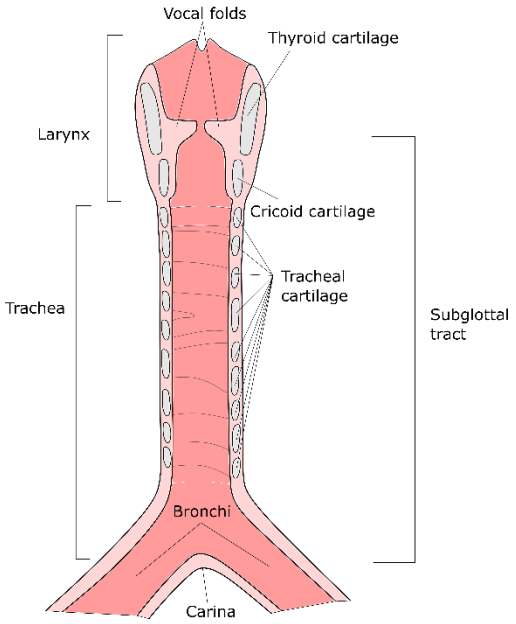


Figure 8. Simplified diagram of the larynx and the subglottal tract.

2.2. Voice instrumentation and parameterization

To better understand the mechanisms of voice production, researchers have designed methods to directly or indirectly visualize and capture the voice, particularly the VF vibrations. This section only describes the techniques, which were used in the papers published as parts of this thesis; other techniques exist but their description is beyond the scope of this thesis.

2.2.1. Laryngoscopic imaging and glottal parameters

2.2.1.1. Standard videolaryngoscopy

Laryngoscopy originates from the need for visually observing the VF vibrations in the larynx. The first instances of laryngoscopy were performed, as early as the middle of the 19th century, by placing a tilted mirror at the back of the mouth, and pointing a light source towards the mirror, in order to have a good visibility of the laryngeal tissues (Moore, 1991; Alberti, 1996; Baken et al., 2000). This simple idea created the basis of modern laryngoscopy, which commonly uses a tube called an endoscope. The endoscope can be either flexible (inserted through the nose, see Figure 9a) or rigid (inserted through the mouth, see Figure 9b), and a camera is commonly connected to it, allowing to record and store the images for later playback and analysis.

Standard videolaryngoscopy can give qualitative insights on the laryngeal configurations and their variations in time, such as the adduction of the VF and the ventricular folds, or the position of the epiglottis and other surrounding tissues. It can also be used in a clinical context to observe the general state of the VFs and other surrounding tissues, and to discover pathologies such as polyps or other lesions (Hirano, 1981). However, the oscillations of the VFs are too fast to be observed and captured using standard videolaryngoscopy.

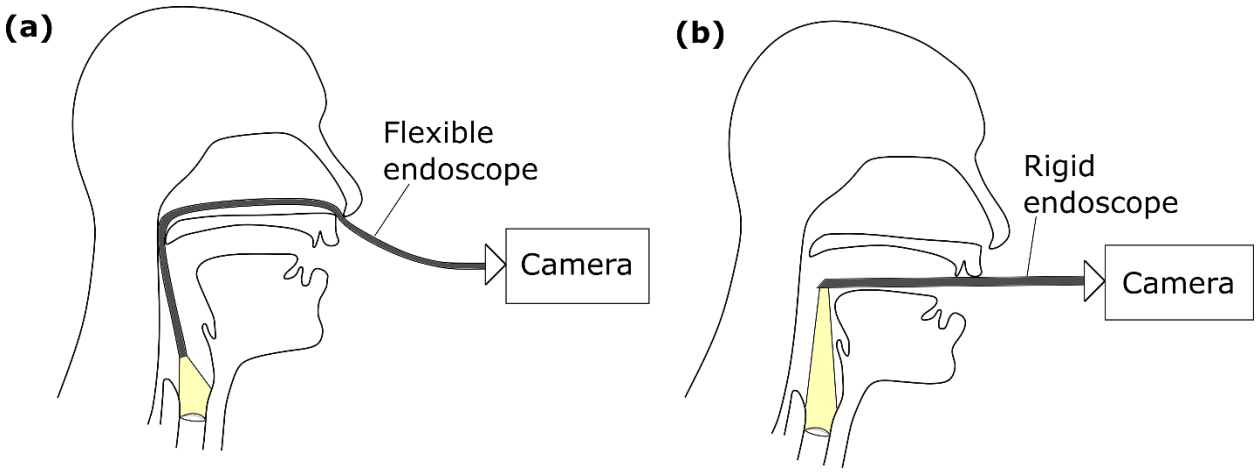


Figure 9. The two main types of endoscope used for laryngoscopic imaging. (a) flexible endoscope through the nasal cavity. (b) Rigid endoscope through the oral cavity.

2.2.1.2. Strobolaryngoscopy

In order to observe the oscillations of the VFs, a technique called strobolaryngoscopy was developed. This technique uses a standard camera with a stroboscopic light source, emitting a series of very short light flashes. The frequency of oscillations of the VFs is measured in real time from the electroglottographic (to be described in section 2.2.2) or the microphone signal, and the frequency of the strobe light is set to be slightly (commonly 1 to 2 Hz) lower than the frequency of oscillations., creating the illusion of a slow oscillation (Kallen, 1932; Moore, 1991; Verikas et al., 2009; Mehta et al., 2010). Strobolaryngoscopy is the most commonly used technique in clinical practice due to its direct, real-time interpretability, and it can be used to estimate parameters related to the VF vibrations (Casiano et al., 1992; Mehta et al., 2010). However, because of the nature of strobolaryngoscopy and its dependence on regular VF vibration with a steady oscillation frequency, one of its major limitations is the impossibility to properly detect irregular phonations.

2.2.1.3. Laryngeal high-speed videoendoscopy

To overcome the limitations of strobolaryngoscopy, researchers can use a high-speed video camera instead of a standard video camera. A high-speed camera can capture images at a much higher rate than the standard (25-50 frames per second – FPS) rate. Technological advances allow cameras to record several tens of thousands of FPS (see, e.g., Herbst, Lohscheller, et al. (2014)). Deliyski, Powell, et al. (2015) found that a frame rate of 8,000 FPS (with a minimum requirement of 4,000 FPS) is capable to observe the VF vibrations with sufficient detail. Laryngeal high-speed videoendoscopy (laryngeal HSV – see Deliyski, Hillman, et al. (2015) for a rationale behind the terminology) allows to observe phenomena which cannot be observed through strobolaryngoscopy, such as irregular vibrations, sudden register changes, or phonation onsets and offsets (Farnsworth, 1940; Eysholdt et al., 1996; Hertegård, 2005). Despite its superior potential for interpretation of the VF vibrations, laryngeal HSV is rarely used in clinical practice, because it does not allow direct, real-time visualization. Since the motion is registered at a very high frame rate, visualization of the frames is very time-consuming and is not suitable in a fast-pace clinical context with many patients.

2.2.1.4. Glottis segmentation and phonovibrography (PVG)

To facilitate the visualization of the VF vibrations captured in laryngeal HSV, image processing techniques can be used to segment the glottis, which means that the glottal area and edges are detected for each frame (Kist et al., 2021). Glottis segmentation allows to extract the time-varying glottal area in pixels (usually referred to as glottal area waveform – GAW) as well as the glottal midline (Kist et al., 2020) and the distance of the left and right VF edge from the glottal midline.

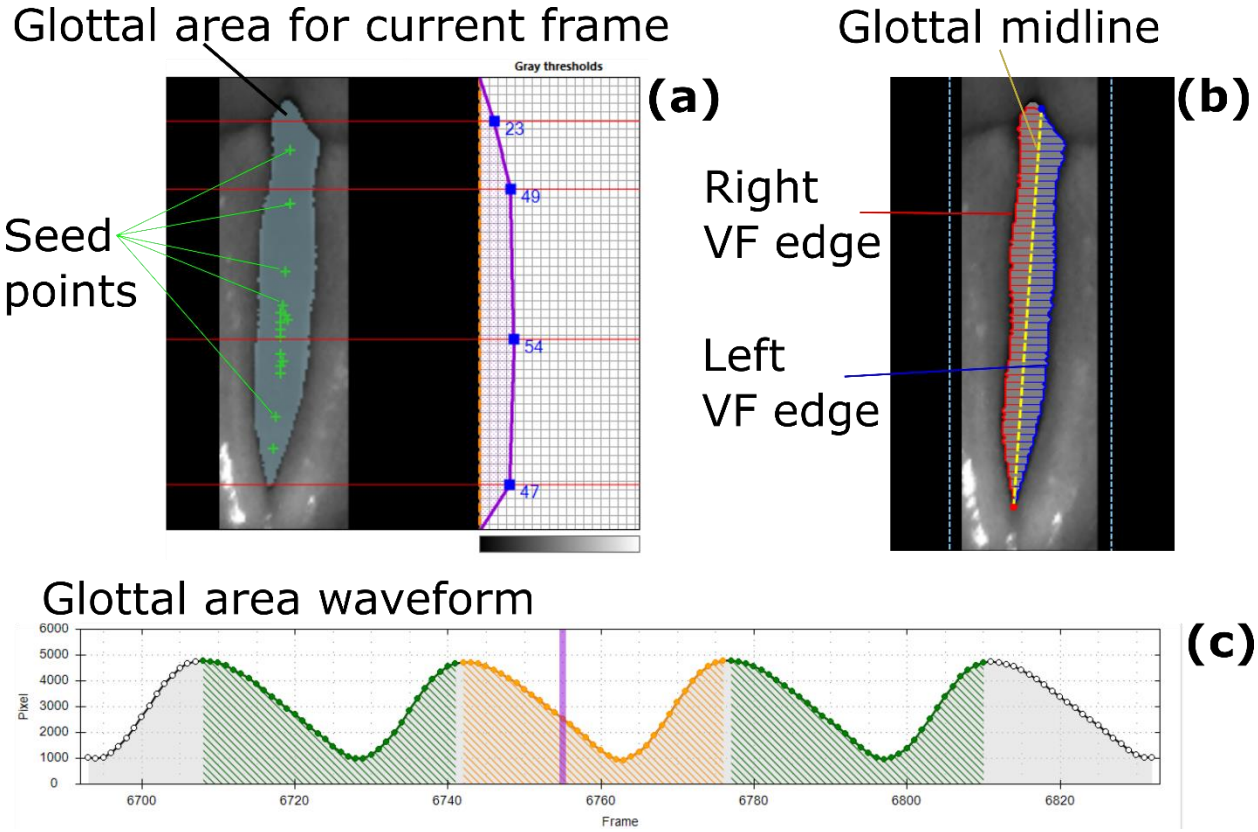


Figure 10. Description of the glottis segmentation from laryngeal HSV data, using Glottis Analysis Tools (Kist et al., 2021). (a) Seed points (green crosses) are placed inside the visible glottis, and the threshold of pixel intensity is adjusted at four different positions along glottal length, so that the detected glottal area (blueish-gray area) matches the visible glottis; (b) Once the segmentation parameters are adjusted to give a satisfactory glottis detection, the glottal midline (yellow dashed line) is automatically detected for each oscillatory cycle, and the distance of the VF edges from this line is coded for each frame; (c) The size of the glottal area (in pixels) is saved for each frame, and displayed as the glottal area waveform (GAW).

Another technique for more practical visualization and quantification of laryngeal HSV is phonovibrography (PVG – Figure 11). This technique consists in detecting the left and right glottal edges and saving their respective distances from the glottal midline (Figure 11a). Then, for each frame, the edge distances along the glottal midline are color-coded (Figure 11b) and concatenated to form the PVG image (Figure 11c). This technique has the benefit of allowing to quickly visualize the motion of the entire glottis, especially regarding opening and closing events (Lohscheller et al., 2008; Döllinger et al., 2009). Other visualization techniques exist, such as glottovibrography (Karakozoglou et al., 2012), but they were not used in the work presented in this thesis, and therefore their description is beyond the scope of this thesis.

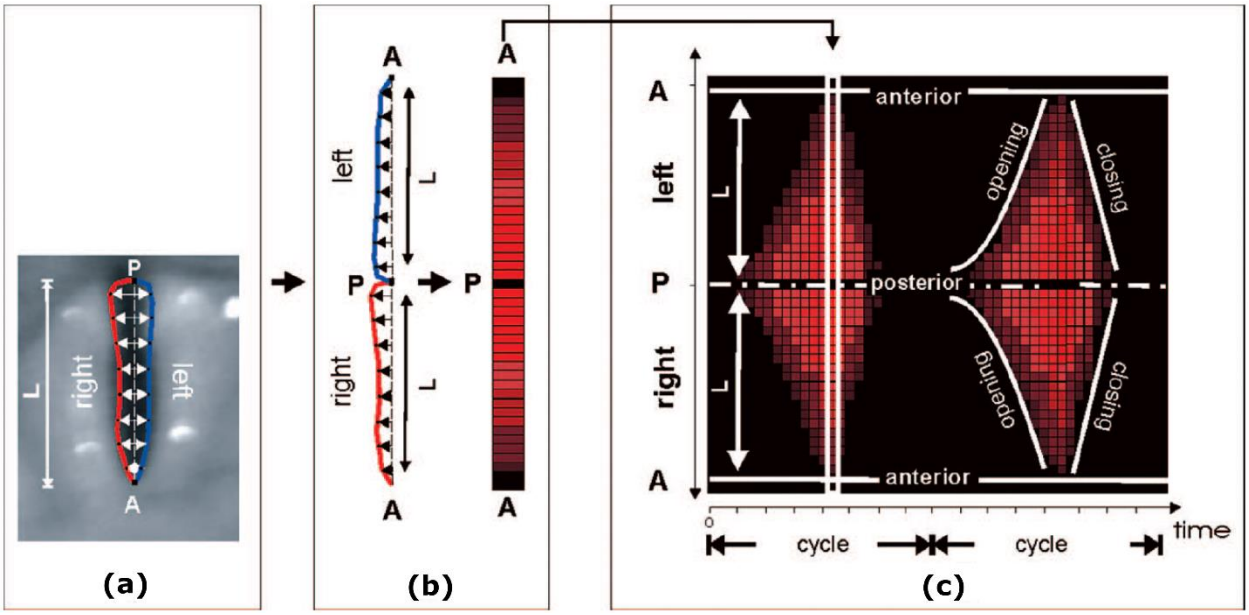


Figure 11. Description of the PVG generation. (a): the glottis is first segmented to detect the glottal edges, for the left and right VFs; (b) the edge contour for the left VF is flipped over the right contour, and the edge distances from the glottal midline are color-coded (here the red color indicates a larger distance, and the black color indicates a smaller distance); (c) the color-coded edge distances are then concatenated for every frame of the HSV. From Lohscheller et al. (2008).

2.2.1.5. Kymography

Kymography is a visualization method used to facilitate the interpretation of HSV images. Digital kymograms (DKG) are images obtained by considering a straight, steady scanline, and concatenating the line images corresponding to this scanline for each HSV frame. An example is shown on Figure 12, where the HSV frame is depicted on the left, showing the straight scanline that was used to extract the line images for every frame, concatenated to form the DKG image, shown on the right.

A technique called videokymography (VKG), developed by Švec et al. (1996), allows the visualization of kymograms in real time. This is done by a modified video camera scanning one horizontal line at a high rate (7,812.5 and 7,200 lines per second for the first and second generation of VKG cameras, respectively – see Qiu et al. (2007) for details about the second generation camera) and displaying the resulting kymographic images at 25 or 30 FPS.

Kymography and VKG have been extensively used to investigate normal and pathological vibrations (Švec et al., 2002; Švec et al., 2007, 2009; Phadke et al., 2017). Notably, VKG has been shown to be useful in assessing the quality of the VF mucosa, through evaluation of the sharpness of the lateral peaks (Jiang et al., 2000; Yamauchi et al., 2015; Kumar et al., 2020). The kymogram image can also be segmented to extract what the author refers to as the kymographic contour, using e.g., VKGanalyzer (Novozámský et al., 2015; Zita et al., 2022).

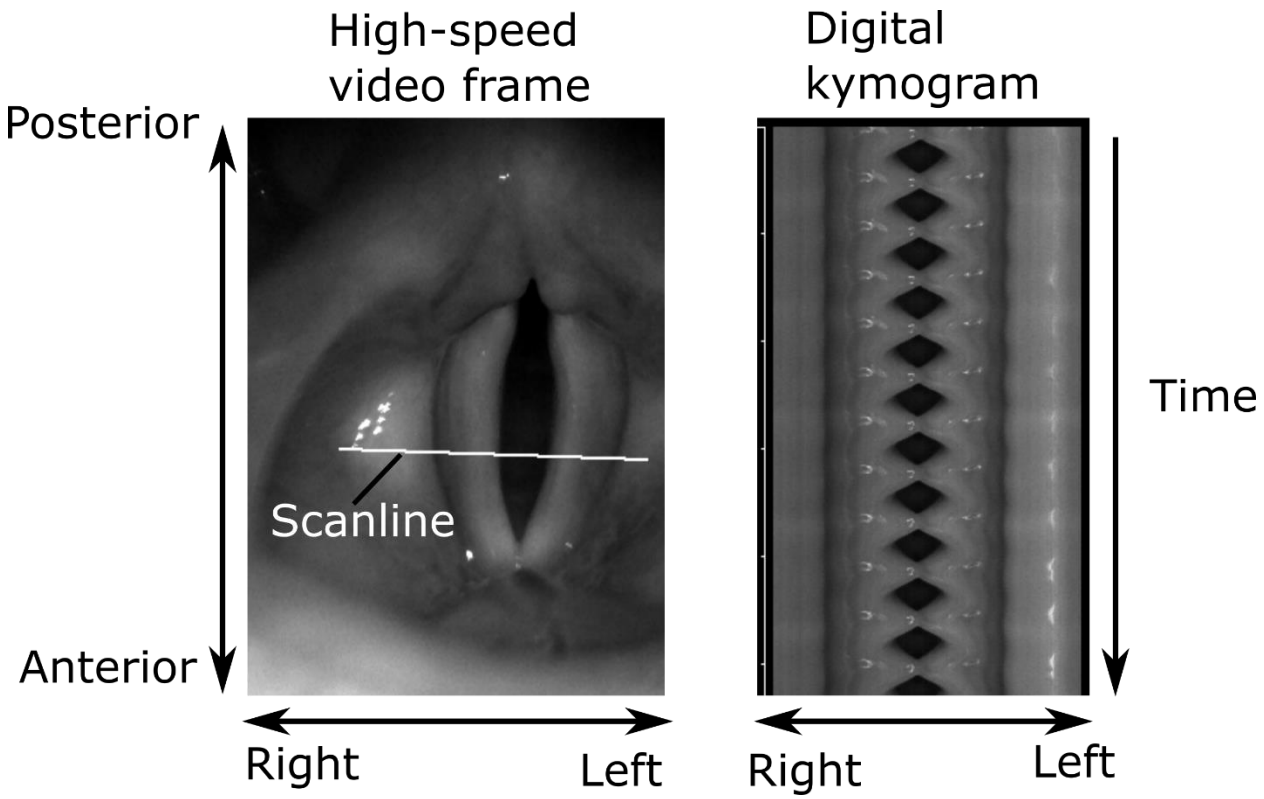


Figure 12. Example of a digital kymogram. The left panel shows a high-speed video frame with the manually drawn scanline. The right panel shows the kymogram image, which concatenates the pixels of the scanline vertically for each subsequent frame.

2.2.1.6. Analysis of the VF vibrations from laryngoscopic images

Segmentation of laryngeal images and kymograms can be used to obtain glottal contours and calculate parameters related to the VF vibrations, which provide a more objective analysis and offer the possibility of comparison and replication. Such parameters are commonly related to the durations of different phases during an oscillation cycle, and their relations to each other. Figure 13 gives a representation of those phases: the oscillation cycle is divided into the closed phase (when the glottis is closed) and the open phase (when the glottis is open), while the open phase is further subdivided into the opening and the closing phases. Here, we define their durations as t_{closed} (closed phase duration), t_{open} (open phase duration), $t_{closing}$ (closing phase duration), $t_{opening}$ (opening phase duration), and t_{cycle} (cycle duration). Their values can be measured from the GAW, the PVG (see section 2.2.1.4), or the kymographic contour (see section 2.2.1.5), and several metrics can be derived (Hirano, 1981). For brevity, only the parameters investigated in paper **III** of this thesis are described here (see e.g. Schlegel et al. (2019) for a more comprehensive list):

- The closed quotient (CQ) is the ratio of the closed phase duration and the cycle duration:

$$CQ = \frac{t_{closed}}{t_{cycle}}. \qquad \text{Eq. (1)}$$

This quotient is sometimes replaced by the open quotient (OQ), which is the ratio of the open phase duration and the cycle duration. The CQ and OQ are substitutable measures, as one can be calculated from the other: OQ = 1 - CQ.

- The speed quotient (SQ) is the ratio of the opening phase and the closing phase durations:

$$SQ = \frac{t_{opening}}{t_{closing}}. \quad \text{Eq. (2)}$$

An alternative expression of the SQ is the speed index (SI), which is the difference between the opening and closing phase durations, divided by the sum of the opening and closing phase durations:

$$SI = \frac{t_{opening} - t_{closing}}{t_{opening} + t_{closing}}, \quad \text{Eq. (3)}$$

- The closing quotient (CgQ) and opening quotient (OgQ) are the ratios of the closing phase duration to the cycle duration, and of the opening phase duration to the cycle duration, respectively:

$$CgQ = \frac{t_{closing}}{t_{cycle}}, \quad \text{Eq. (4)}$$

$$OgQ = \frac{t_{opening}}{t_{cycle}}, \quad \text{Eq. (5)}$$

- The amplitude quotient (AQ) is the ratio of the amplitude of vibration (see Figure 13) and the maximum declination rate of the considered signal, which is defined as the absolute value of the minimum derivative of the signal (analog to the maximum closing speed) (Titze, 2006c). This quotient can also be normalized by the cycle duration, in which case it is commonly called the normalized amplitude quotient (NAQ) (Alku et al., 2002):

$$NAQ = \frac{amplitude}{maximum\ declination\ rate \cdot t_{cycle}}. \quad \text{Eq. (6)}$$

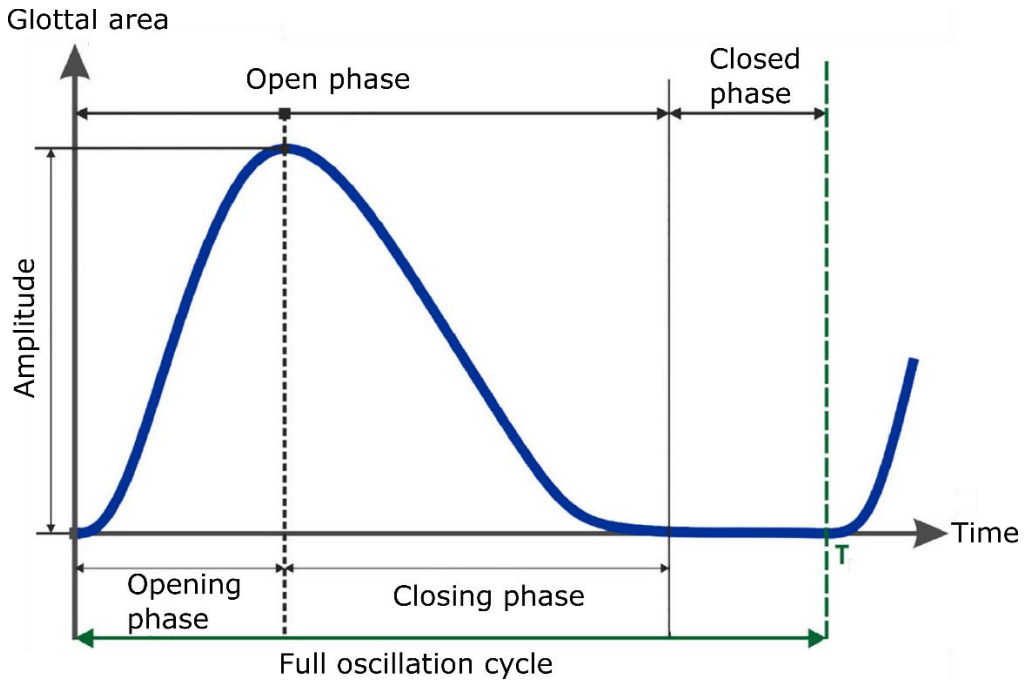


Figure 13. Description of the phases within an oscillation cycle, based on the glottal area waveform. Modified from Schlegel et al. (2019).

Finally, the kymographic contour can be used to estimate the vertical phase differences between the lower and upper margins of the VFs. The technique used to this purpose will be described in section 2.3.3.

2.2.2. Electroglottography

Electroglottography (EGG) is a non-invasive method where two electrodes measure the electrical conductance across the larynx, as represented on Figure 14a (Fabre, 1957; Baken, 1992; Herbst, 2019). Because of the differences in conductance between air and the vocal fold tissues, it is generally assumed that the EGG signal is related to the vocal fold contact area (VFCA) (Scherer et al., 1988; Titze, 1989, 1990; Hampala et al., 2016), making it a useful tool for visualizing the VF vibrations. It is especially used to investigate the contacting (when the VFs increase contact between them) and decontacting (when the VFs decrease contact between them) instants. An idealized EGG waveform is depicted in Figure 14b: the signal commonly contains a low, flat part which corresponds to the open phase (before instant A on Figure 14b), then increases (as the lower margins of the VFs suddenly come in contact – instant A on Figure 14b) to reach a maximum value (when the contact is strongest – instant C on Figure 14b), then decreases as the VFs start decontacting, opening the glottis (instant D on Figure 14b), to finally reach a low value as the glottis is completely open (instant F on Figure 14b). The presence of a "knee" (i.e., a prominent convexity along the decontacting slope – instant E in Figure 14b) seems typical for men's voices, while women's voices appear to exhibit a smoother EGG amplitude decrease (Patel et al., 2021).

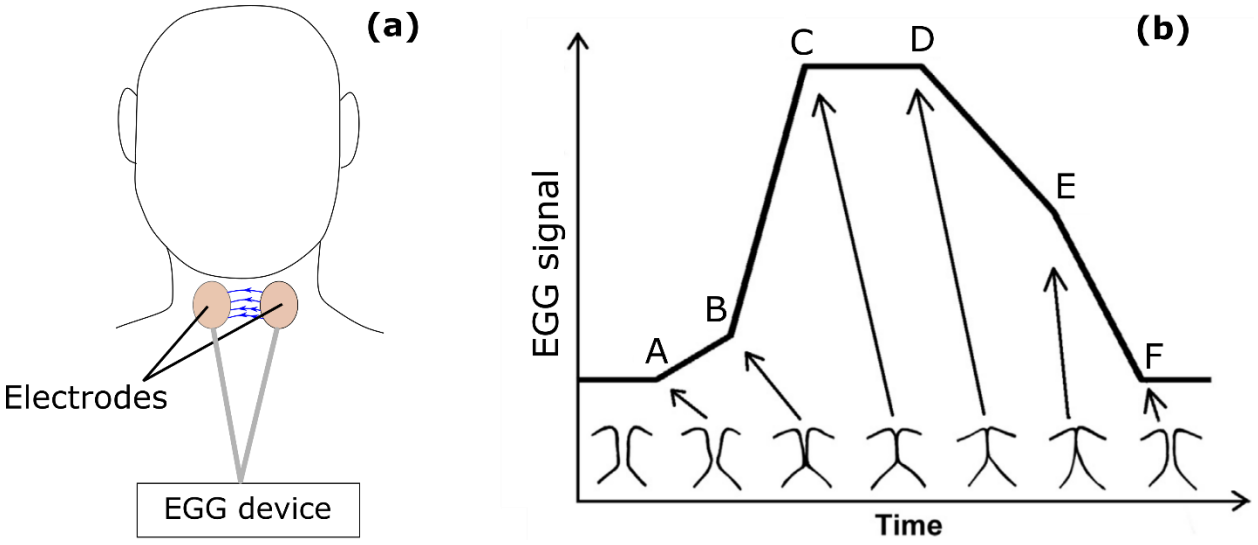


Figure 14. Description of the electroglottography (EGG) technique. (a) the two electrodes are placed across the larynx, at the position of the glottis, and connected to the EGG device; (b) idealized schematic interpretation of the EGG waveform: A – initial contact of the lower VF margin; B – initial contact of the upper VF margin; C – maximum VF contact reached; D – decontacting phase initiated by separation of the lower VF margin; E – upper margin starts to separate; F – the glottis is open, the contact area is minimal. Modified by the author from Hampala et al. (2016) and Baken (1992).

The approximate instants of contacting and decontacting can be estimated using different methods, such as considering a threshold value on the normalized EGG waveform, or using the maximum and minimum values of the first time derivative (Henrich Bernardoni et al., 2004; Herbst, 2019). Those instants are commonly used to measure the contact phase duration $t_{contact}$ and derive EGG the contact quotient CQ_{EGG}:

$$CQ_{EGG} = \frac{t_{contact}}{t_{cycle}}. \quad \text{Eq. (7)}$$

It is sometimes expressed as the EGG open quotient $OQ_{EGG} = 1 - CQ_{EGG}$. Ternström (2019) recently proposed another method to measure the contact quotient by integrating the normalized EGG waveform, avoiding the need of determining the instants of glottis opening and closing.

The average amplitude of the EGG signal also depends on the vertical position of the electrodes on the larynx, and it will be maximum when the electrodes are placed at the level of the VFs. Slow vertical motions of the larynx can occur in speech in singing, and therefore a drop in the average amplitude could either suggest a decrease in the VF contact strength, or a vertical larynx motion, or both. To overcome this limitation of EGG, some modern EGG devices also include multiple-channel electrodes in order to detect vertical motions of the larynx and estimate its relative vertical position with regards to the electrodes (Rothenberg, 1992; Kob et al., 2009).

2.2.3. Acoustic recording and characteristics of the voice

Acoustic recordings of the voice sound can be used for perceptual evaluation (for example to understand how voice registers are perceived, as described in paper **III** of this thesis – see also section 2.4 regarding voice registers). They are also used for acoustic analysis to investigate acoustic properties of the voice sound, such as its spectral content (i.e., the energy distribution as a function of frequency), its fundamental frequency and intensity, their variations and perturbations in time, etc. This section first describes the basic acoustic features of the voice sound, and then describes parameters that can be measured through acoustic analysis.

2.2.3.1. Source-Filter theory of voice production

The basic acoustic theory of voice production is called the "source-filter" theory (Fant, 1960). This theory describes the voice production as two systems: the source (airflow modulated by the VF oscillations – see Figure 15a-b) and the filter (vocal tract resonances – see Figure 15c-d). The frequency spectrum of the source glottal flow exhibits several harmonics (peaks in the spectrum – sometimes referred to as "partials"), which are multiples of the lowest harmonic (Figure 15b). In regular voices, the frequency of the lowest harmonic normally corresponds to the fundamental frequency of oscillation f_o (Titze et al., 2015). As mentioned in section 2.1.1.2, the supraglottal tract presents acoustic resonances which act as a filter on the source sound (Figure 15d), and the resonance frequencies depend on the supraglottal tract geometry (Figure 15c). The output voice sound is the result of the linear combination of the source and the filter (Figure 15e-f), and the local maxima in the envelope of the spectrum are called formants (Titze et al., 2015).

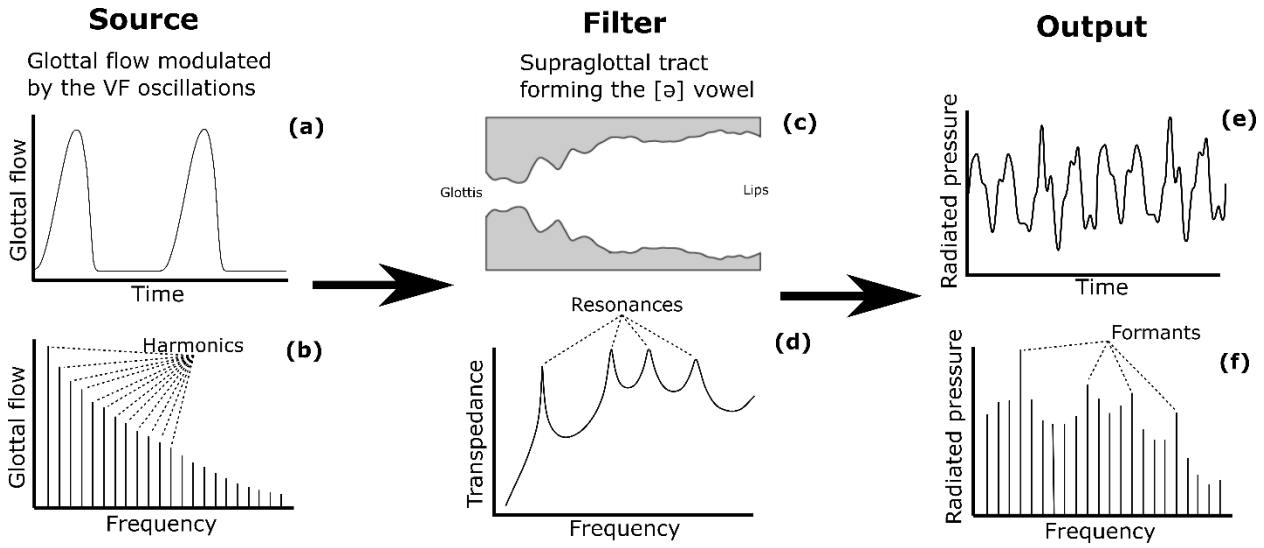


Figure 15. Description of the source-filter theory of voice production. (a) Source: glottal flow waveform generated by the VF oscillations; (b) Spectrum of the glottal flow; (c) Filter: geometry of the supraglottal tract; (d) Transfer function of the supraglottal tract corresponding to the geometry in (c); (e) Sound output: radiated pressure time signal; (f) Spectrum of the radiated pressure. Based on Wolfe et al. (2016), Fig. 2, modified by the author.

2.2.3.2. Acoustic analysis of the voice sound

This subsection describes two essential acoustic characteristics of voice: the fundamental frequency of oscillation f_o and the sound pressure level (SPL); additionally, the spectral parameters used in the papers of this thesis are described.

Fundamental frequency of oscillation f_o

The frequency of oscillation f_o, measured in the units of hertz (Hz), defines the number of oscillation cycles per second, and it can be measured from the voice sound. It is often associated (especially in singing voice) with the term "pitch", which is a perceptual term used to describe whether a sound is perceived to be "high" or "low". Within a musical context, this term is commonly used to refer to musical notes. In Western occidental music, the notes are called (in English) A, B, C, … G, and they are paired with a number. The number is related to a particular octave of the note: going from one octave to a higher one means that f_o is doubled, and going to the lower octave means that f_o is halved. For example, the note A4 is commonly used for instrument tuning and is standardized to correspond to f_o = 440 Hz. Consequently, the notes A3 and A5 have the f_o of 220 and 880 Hz, respectively (see the webpages from Wolfe (2001) and Botros (2001) for the correspondences between all notes and their f_o value).

Sound pressure level (SPL)

Any sound, including the voice sound, is also characterized by how loud it is perceived. There exist several objective measures associated with the perceptual concept of "loudness", but the SPL is the most commonly used for the voice sound, both in clinical and research contexts. The value of the SPL L_p is measured using the following equation (Švec et al., 2018):

$$L_\mathrm{p} = 20 \log_{10} \frac{p}{p_0}, \qquad \text{Eq. (8)}$$

where p is the measured sound pressure and p_0 is the reference sound pressure which equals 20 µPa in air. As microphones usually deliver an electric signal which is amplified by an arbitrary gain, the values of the microphone signal are usually not inherently scaled with any reference. Therefore, to calculate the SPL, the microphone signal needs to be calibrated so that its values are scaled towards a reference unit, usually in the units of pascal (Pa). The SPL is commonly calculated as the average over a period of time, and a frequency weighting can be applied to approximate the sensitivity of the human ear. The extensive description of the SPL calculation and microphone signal calibration is beyond the scope of this thesis, but readers are referred to the tutorial by Švec et al. (2018) for such a description. The SPL has clinical and research relevance, as it has been shown to be positively correlated with the logarithm of the mean subglottal pressure (Bouhuys et al., 1968; Sundberg et al., 1999; Björklund et al., 2016; Sundberg, 2018), which is usually associated with vocal effort.

Voice sound spectral characteristics

The spectral content of the voice sound describes its "timbre", which is a perceptual notion. How a voice sound is perceived can be correlated with spectral features in the energy distribution along frequency. For example, a voice that sounds bright, rich, or strong, contains more energy in the high-frequency part of the spectrum, compared to a voice that sounds dark, mellow, or dull. Differences in timbre can be the result of adjustments at the laryngeal level (e.g. VF adduction) or changes in the supraglottal tract geometry, but for brevity only the former is described in this section. Differences resulting from laryngeal adjustments can be studied from the glottal source sound, which is obtained by the so-called inverse filtering technique, either from the microphone signal or from the oral flow signal captured using a special flow mask (Rothenberg, 1973; Baken et al., 2000; Alku, 2011). The process of inverse filtering generally involves removing formants frequencies from the voice sound to obtain the sound source signal which resembles the one depicted in Figure 15a. The sound source signal has notably been used in previous studies to measure the difference in the levels of the first and second harmonic L_{H1}-L_{H2}, measured from the inverse-filtered voice signal, to assess the spectral characteristics associated with different phonation types (Sundberg et al., 1999; Sundberg et al., 2001; Kreiman et al., 2007; Sundberg, 2022). Particularly, Sundberg (2022) noted that the values of L_{H1}-L_{H2} were lowest for pressed phonations and highest for breathy phonations, and that L_{H1}-L_{H2} decreased when the glottal flow amplitude was increasing. Zhang (2016), through a three-dimensional numerical model of the VFs, also reported lower L_{H1}-L_{H2} values when increasing the VF thickness (simulated by the contraction of the TA muscle – recall Figure 6).

To assess the more global differences in the voice timbre, researchers have measured and compared the energy in the low- and high-frequency regions of the voice sound spectrum. Particularly, the spectral balance (SB) is a measure of the difference between the sound levels within the high-frequency and low-frequency bands. The definition of the low-frequency and high-frequency bands seems to differ among studies, however this thesis uses the method Collyer et al. (2007), which defines the low-frequency band from 0 to 2000 Hz, and the high-frequency band from 2000 to 4000 Hz. The SB can reflect the overall timbral quality of the voice sound, and it was found to be positively correlated with the SPL, indicating that increasing the SPL is related to increasing the energy in the high-frequency bands and flattening the spectral slope (Collyer et al., 2007, 2009).

2.3.Myoelastic aerodynamic (MEAD) theory of voice production

In practice, the source-filter model does not attempt to explain the mechanisms of VF self-oscillations, as well as various nonlinear dynamic phenomena of voice. For this purpose, the MEAD theory of voice production was explicitly formulated by Van den Berg (1958) and refined throughout the years by e.g., Titze (1980, 1988b, 2006b) and Švec et al. (2021). This theory opposed the so-called neuro-chronaxic theory, which assumed that the VF oscillation was caused by very fast contractions of the TA muscle (Husson, 1950).

The MEAD theory states that the VFs are driven by a combination of aerodynamic forces and of the elastic properties of the VF tissue. The aerodynamic forces are generated by the airflow created from an increase in the mean subglottal pressure, and the elastic properties of the VF tissue can be modified through the activity of laryngeal muscles. The elastic properties of the VFs play a role in the frequency of VF oscillation f_o, which is described in the following subsection.

2.3.1. Control of f_o

To understand how the properties of the VFs affect their oscillation frequency, the VFs can be modelled as strings or as simple lumped element (mass-spring-damping) systems. Notably, Titze et al. (1988) and Verdolini et al. (1994) applied the string equation, which relates the frequency of oscillation f_o to the VF length L, the VF tissue density ρ, and the VF tensile stress σ:

$$f_o = \frac{1}{2L}\sqrt{\frac{\sigma}{\rho}}. \qquad \text{Eq. (9)}$$

This equation, as an approximation, helps to understand that f_o is inversely proportional to the length of the VFs, and proportional to the square root of the stress σ.

Increasing f_o can be achieved by increasing the stress σ in the VF tissues, which can be done by two different ways:

- Activating the TA muscle, which bulges the VFs and actively tenses the muscle layer (see Figure 6).
- Activating the CT muscle, which pulls the thyroid cartilage forward and rotates it with respect to the cricoid cartilage (see Figure 3). This elongates and tenses the VFs.

The TA and CT muscles are antagonists, which means that the action of one counteracts the action of the other. In practice, varying f_o often results from a combination of the activity of both muscles. Figure 16 demonstrates the theoretical effect on f_o from possible combinations of the CT and TA muscle activities (Titze et al., 1988; Titze et al., 1989; Titze, 2000b). From Figure 16, it is clear that the action of each muscle has a different impact on the variation of f_o. In particular, two possible paths to increase f_o are shown as black arrows on Figure 16. The path which starts lower than the diagonal (straight arrows) represents a greater activity in the TA muscle than in the CT muscle. However, once the TA muscle reaches its maximum activity, no more increase in f_o can be achieved in this direction. The solution to further increase f_o is usually to abruptly disengage the TA muscle. This is assumed to create a voice break and a sudden jump to a higher f_o, as indicated by the second straight arrow on Figure 16. This kind of jump is commonly referred to as a frequency jump, or sometimes pitch jump. It is typically observed during register transitions (registers will be described in more details in section 2.4). Frequency jumps can be cultivated in certain styles such as yodeling, but singers are more commonly trained to avoid them and to execute a

smooth f_o increase, without any perceivable transition. This can be performed by following the other path depicted by the curved arrow in Figure 16: the activity of the CT muscle needs to be gradually greater while gradually disengaging the TA muscle (Titze, 2000b).

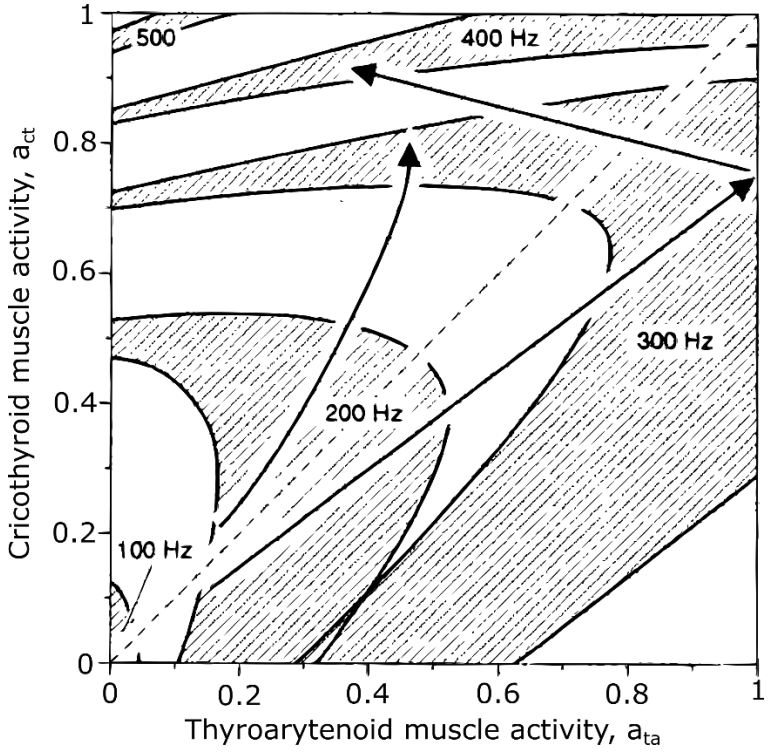


Figure 16. Muscle activation plot (MAP) indicating theoretical effects of CT muscle activation (a_{CT}) and TA activation (a_{TA}) on f_o. The two straight arrows and the curved arrow indicate different paths to increase f_o. Modified from Titze (2000b), Fig. 10.11.

2.3.2. Acoustic interactions with the supraglottal and subglottal tract

Generally, the MEAD theory suggests that the VF vibrations are influenced by the oscillating acoustic pressures below and above the VFs (Herbst et al., 2023). Titze (2008) proposed to separate the acoustic interactions between the VF vibration and the supraglottal or subglottal pressures into two groups, denoted as level 1 and level 2 interactions. Level 1 interaction occurs when the supraglottal or subglottal tract acoustic properties affect the glottal flow waveform, for example when the flow pulse is skewed from the effect of the supraglottal tract inertance (Rothenberg, 1981; Titze, 2006c). Level 2 interaction occurs when the supraglottal or subglottal tract acoustic properties affect the vibrational pattern of the VFs, the most notable example of such interaction being the occurrence of frequency jumps when f_o is in the vicinity of the first supraglottal tract formant frequency. This has been demonstrated both in vivo and in numerical simulations (Titze, 2008; Titze et al., 2008; Wade et al., 2016, 2017; Murtola et al., 2018). An example of such frequency jumps is shown in Figure 17: the lowest trace in the spectrogram shows the value of f_o and the horizontal line displays the estimated frequency of the first supraglottal formant (here denoted as F1). Small black arrows indicate sudden frequency jumps, and the arrows #1 and #3 clearly indicate that the value of f_o suddenly jumped from above to below F1 (arrow #1) and from below to above F1 (arrow #3). This suggests that the first supraglottal resonance pushed the VF vibrations away from vibrating at the frequency of the first supraglottal tract resonance.

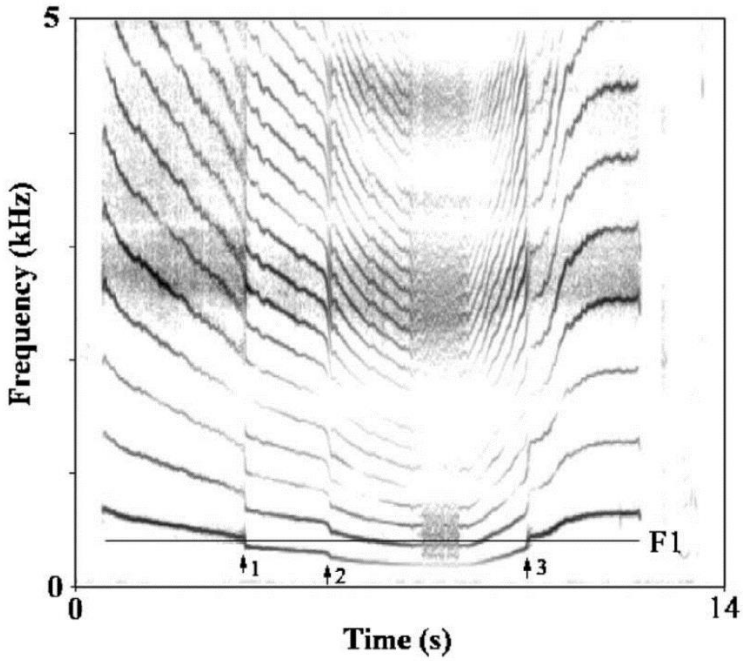


Figure 17. Spectrogram of a downward and upward glissando, showing occurrences of frequency jumps (see vertical arrows). The horizontal line indicates the estimated frequency of the first formant (F1). Reprinted from Titze et al. (2008), Fig. 5.

Besides of level 2 VF interactions with the supraglottal tract, similar effects have also been hypothesized by Titze (1988a) to occur from VF interactions with the subglottal tract. It has been experimentally (with a physical model of the VFs) demonstrated by Zhang et al. (2006) that the subglottal resonances can also entrain the frequency of oscillation of the VFs. The influence of the subglottal resonances on the VF oscillations is further investigated with excised larynx experiments (to be described in section 2.5) in papers **I** and **II** in this thesis, which are related to level 1 and level 2 interactions with the subglottal tract.

The inertance of the supraglottal tract has been shown to play a crucial role for level 2 interactions and consequently also for driving the VF oscillations: the inertia of the air in the supraglottal tract generates a positive supraglottal pressure during the glottal opening and a negative supraglottal pressure during glottal closing, which helps to transfer the aerodynamic energy into the VFs for self-sustained oscillations (Rothenberg, 1981; Titze, 2000d; Švec et al., 2021). This phenomenon was demonstrated already in the numerical one-mass model by Flanagan et al. (1968), which was unable to vibrate without a supraglottal tract. However, excised larynx experiments showed that VFs can exhibit self-sustained oscillations even without a supraglottal tract, which suggests that there exists another mechanism that drives the VFs and provides a sufficient energy transfer from the airflow. This mechanism is related to the nature of the VF vibration pattern, particularly with the vertical phase differences related to the mucosal waves that propagate along the VF surface. This mechanism of self-oscillation is described in the next section.

2.3.3. Vocal fold mucosal waves and vertical phase differences

For the VFs to self-sustain oscillations, the energy transfer between the airflow and the VF tissue must be sufficient to overcome the natural damping caused by the tissue viscosity. Titze (1988b) argues that, without the inertance of a supraglottal tract, the energy transfer is not sufficient to sustain oscillation if the shape of the VF medial surface does not change during the VF oscillatory phases, as depicted in Figure 18a. Instead, the shape of the VF medial surface must exhibit a change between the opening and closing

phase, such that it is convergent during the opening phase and divergent during the closing phase, as depicted in Figure 18b. This pattern of vibration is generated by a vertical phase difference between the vibration of the lower margin and the upper margin of the VFs. Ishizaka et al. (1972) demonstrated this pattern of vibration with a two-mass model of the VFs, which was capable of sustaining oscillations even without a supraglottal tract. The vertical phase differences create surface waves that propagate along the medial and top surface of the VFs, which are called mucosal waves.

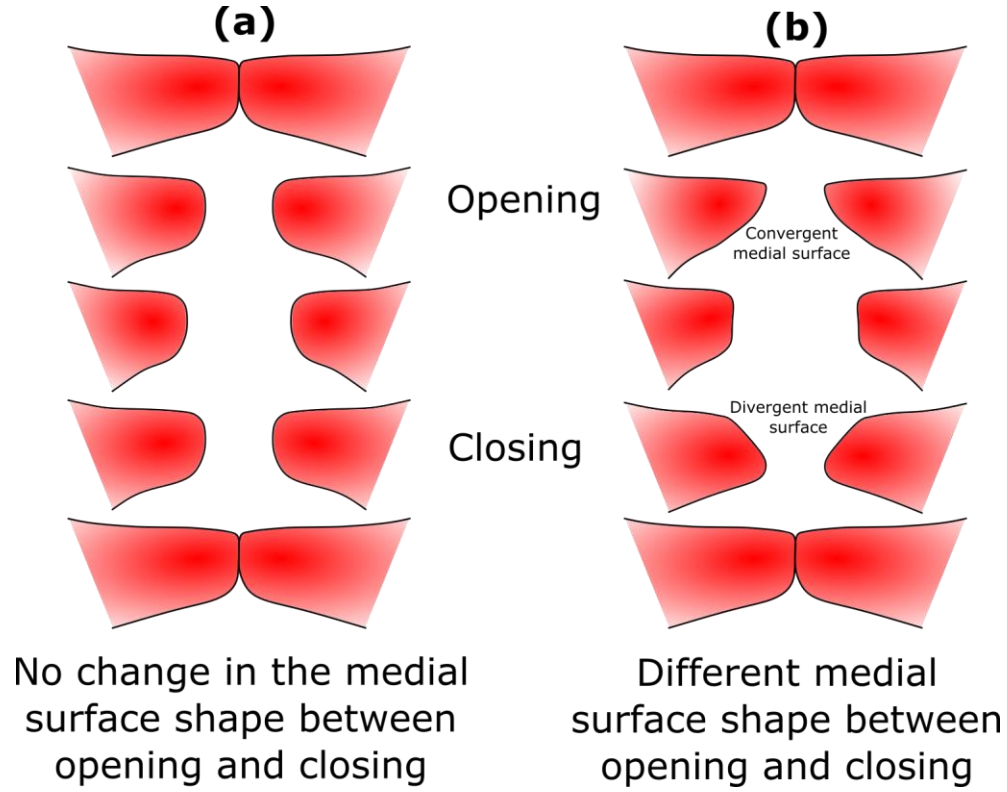


Figure 18. Oscillation cycle of the VFs. (a) Constant shape of the VF medial surface: the VFs vibrate as a whole, without any vertical phase difference; (b) the VFs vibrate with a vertical phase difference which creates a convergent shape during the opening phase and a divergent shape during the closing phase.

The mucosal wave characteristics and behavior can be a useful indicator of the presence of voice disorders and has been extensively studied. Titze et al. (1993), Hirano et al. (1981) and Baer (1975) measured the mucosal wave velocity and reported values around 0.1 to 0.5 m/s for the superior surface and around 1 m/s for the medial surface. Other measurements have been made in vivo to estimate the phase difference between the lower and upper margin of the VFs, which is a direct inverse correlate of the mucosal wave velocity. Jiang et al. (2000) proposed a way to estimate the vertical phase difference using kymographic images of the VF vibration. The method consists in approximating the motions of the lower and upper margins of the VFs as sinusoids, and considering that the kymographic contour is determined by the upper margin during the opening phase and by the lower margin during the closing phase. Both upper and lower margin movements during the open phase are then independently fitted to sinusoidal curves (see Figure 19), and the phase difference is estimated as the difference between the initial phases of the lower and upper margins. A more recent description of this method can be found in Jiang et al. (2008).

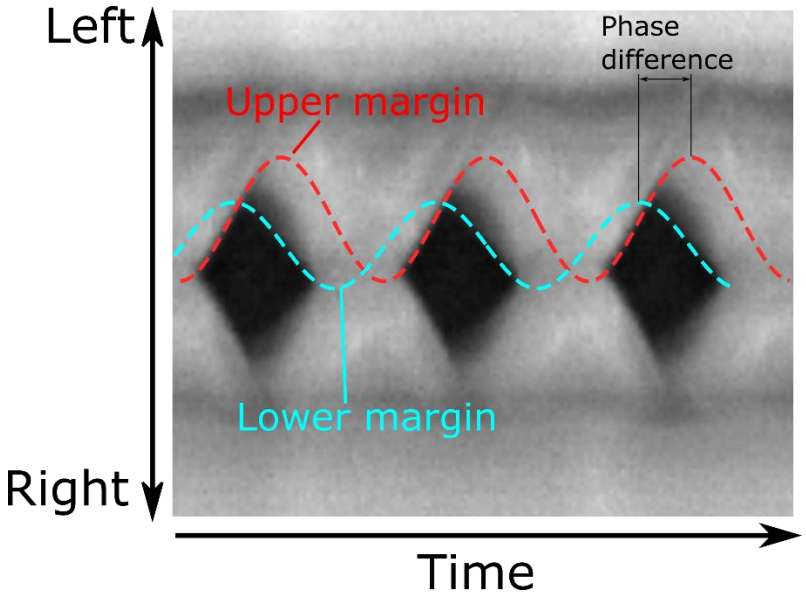


Figure 19. Kymogram of a steady phonation, showing sinusoid curves fitted to the upper (red curve) and lower (teal curve) margins for the right VF.

The mucosal wave behavior can also be qualitatively assessed by looking at the sharpness of the lateral peaks on the kymographic contours (Hiroto, 1966; Sundberg et al., 2001; Švec et al., 2007; Krausert et al., 2011; Phadke et al., 2017; Kumar et al., 2020): the sharper the lateral peaks, the higher the vertical phase difference between the lower and upper margin. Generally speaking, a higher vertical phase difference (and therefore, a lower mucosal wave velocity) indicates a high pliability of the VF mucosa, which is considered as an indicator that the mucosa is rather healthy and not damaged. The mucosal wave behavior has also been investigated in modelling studies, which will be described later in section 2.6.

2.3.4. Eigenmodes of vibration

The oscillating motion of the VFs can be decomposed into a series of natural modes of vibration, which are commonly referred to as eigenmodes, while each eigenmode is associated with a natural frequency of oscillation, called eigenfrequency (see e.g., Kinsler et al. (2000) for a general description of modes of vibration). The VFs possess an infinite number of possible eigenmodes, however it was shown that in regular phonation, two eigenmodes are often sufficient to explain a high proportion of the vibration (Berry et al., 1994; Berry & Titze, 1996; Berry, 2001).

During regular phonation (sometimes referred to as modal phonation), the vibration is periodic (with some slight degree of irregularity in the period and the amplitude). In this case, the principal eigenmodes of vibration are synchronized: their respective eigenfrequencies are very close together. This phenomenon is referred to as eigenmode entrainment by Berry (2001), and it is a nonlinear phenomenon facilitated by e.g., vocal fold collision and aerodynamic forces. A desynchronization of the eigenmodes (i.e., the eigenfrequencies are farther away from one another) causes the phonation to be no longer regular, evoking vocal instabilities (Herzel et al., 1994). This phenomenon offers another mechanism to explain the occurrences of sudden frequency jumps described in section 2.3.1: a small variation of input parameters such as the VF tension could cause a desynchronization of the eigenmodes, abruptly followed by an eigenmode entrainment to another frequency.

In the field of nonlinear dynamics, a system abruptly switching from one regime to another when input parameters are smoothly varied is called a bifurcation. Bifurcations occur within ranges of parameters where two or more dynamic regimes coexist (Berry, Herzel, et al., 1996). Notably, frequency jumps are often described as bifurcations, where the oscillations suddenly change their frequency to another.

2.4.Registers

The concept of voice registers is a controversial topic in the voice research field, and there seems to be no universal consensus on the definition and the number of registers (Mörner et al. (1963) listed more than a hundred different terms to refer to registers). The use of the term "register" was originally borrowed from organ terminology (Merkel (1863), as cited in Large (1972)): the differences in timbre between regions of notes sung on each side of the passagio (transition tones where a register change typically occurs) were analogous to different sets of organ pipes called registers, and the term was then used also for singing (Alderson, 1979).

The timbral differences have therefore been used as a way to define and describe registers. Other approaches include using proprioceptive sensations, for example the sensation of "resonance" in certain parts of the body, giving rise to the terms chest register and head register (Stark, 1999), or the differences in the vibratory pattern of the VFs (Henrich Bernardoni et al., 2003; Henrich Bernardoni, 2006; Roubeau et al., 2009). This latter approach was originally formulated by Manuel Garcia (1847), who described registers as a series of "consecutive and homogeneous tones […] produced by the same mechanical principle, and whose nature differs essentially from another series of tones equally consecutive and homogeneous produced by another mechanical principle". It is generally recognized that there are four main registers:

- Vocal fry (or M0, pulse, strohbass): describes pulse-like VF vibrations, with a low f_o, typically below 70 Hz (Titze, 1988a).
- Chest (or M1, modal): describes the main register used for speech and the lower singing range.
- Head (or M2, falsetto, loft): describes the register used for the upper singing range.
- Whistle (or M3, flute, flageolet): describes the register used for the extreme high end of the singing range. The sound is very flutey, like a whistle (Švec et al., 2008; Garnier et al., 2012).

Other approaches exist, based on pedagogic experience, that define more than four registers (Sadolin, 2000) or a mix between two registers (Castellengo et al., 2007), but the description of those approaches is beyond the scope of this thesis.

2.4.1. Laryngeal mechanisms

Roubeau et al. (1987) were among the first to use the term "mechanism", particularly to refer to the chest (mechanism I) and head (mechanism II) registers. Later, the term "laryngeal mechanism" (and the associated terms M0, M1, M2 and M3) was proposed by Henrich Bernardoni et al. (2003). This terminology was used to suggest that there are distinct vibratory regimes of the VFs, characterized by an observable transition between them. This transition can take the form of a frequency jump or an abrupt change in the amplitude of the EGG signal (recall section 2.2.2) or in the EGG contact quotient (Henrich Bernardoni et al., 2003; Henrich Bernardoni et al., 2004, 2005; Roubeau et al., 2009). Roubeau et al. (2009) suggest that there can be great timbral and intensity variation in the sounds produced within one laryngeal mechanism, and that even though the notions of register and of laryngeal mechanism overlap,

they should be treated as different. They also suggest that each laryngeal mechanism is somewhat limited to a certain frequency range, even though two neighboring laryngeal mechanisms can have partially overlapping frequency ranges.

2.4.2. Characteristics of chest and head registers

As previously mentioned, one possible way of recognizing registers is to use one's own kinesthetic sensations, the development of which is commonly done through singing training. Singing teachers usually provide feedback based on the sound that is produced during singing practice, considering that their perception of this sound is shaped by their own practice and pedagogic experience. Nevertheless, the characteristics of registers have been studied through various methods, which are described in this section. For brevity, this section will only deal with the investigation of chest and head registers.

The laryngeal differences between chest and head registers have been mostly revealed by early electromyographic studies which investigated the activity of the intrinsic and extrinsic laryngeal muscles during phonation (Hirano et al., 1970; Baer et al., 1976). In particular, those studies demonstrated that the TA muscle activity is high in chest register and falls when shifting from chest to head register. This was further supported by more recent studies measuring the activity of the TA and CT muscles (Kochis-Jennings et al., 2012; Kochis-Jennings et al., 2014). The theoretical framework of Titze (2000b) provides possible explanations for the need to disengage the TA muscle as f_o increases, as mentioned in section 2.3.1.

Physiologically, besides of increasing the tensile stress in the body of the VFs, the contraction of the TA muscle bulges the VFs (recall Figure 6), which allows for a higher membranous adduction (Herbst et al., 2011; Herbst & Švec, 2014). This can increase the duration and strength of the VF contact, which in turn generates an increase of the energy in the high-frequency harmonics, typically observed through a flatter spectral slope in chest register (Colton, 1972). In practice, this is demonstrated by the higher closed quotient values measured for chest register phonations (Sundberg et al., 2001; Roubeau et al., 2009; Echternach, Burk, Köberlein, Herbst, et al., 2017; Echternach, Burk, Köberlein, Selamtzis, et al., 2017). Values of L_{H1}-L_{H2} were also found to be lower for chest register phonations (Sundberg et al., 2001). In addition, the contraction of the TA muscle increases the thickness of the VFs (Hirano, 1974; Zhang, 2016), which can increase the vertical phase differences (recall section 2.3.3) between the lower and upper margins of the VFs. Thicker VFs can also increase the amplitude of vibration of the lower margin, which increases the glottal closing speed.

2.5.Excised larynx experiments

Despite its numerous benefits, in vivo investigation of voice production only offers limited possibilities due to ethical and technical constraints. To obtain deeper insights into the physiological and biomechanical aspects of voice production mechanisms, physical models of the larynx can be used, which offer better parameter control and more analysis possibilities than in vivo. Those models can be either synthetic (made from synthetic material such as silicone) or larynges excised after death. Both types of models have been widely used in research, but this thesis will only focus on the description of excised larynx experimental methods, which were utilized in papers **I** and **II**.

2.5.1. Brief history of excised larynx experimental setups

Early experiments from Ferrein (1741) with excised larynges were important to understand the most basic mechanisms of voice production, such as the fact that the vibration of the VFs was the source of the voice sound and that the tension applied to the VF varied f_o, although Ferrein believed that the VFs were producing sound in the same manner as strings which drive the air (Van den Berg, 1958). Later technological advances allowed researchers to perform more thorough measurements on excised larynges and to design experimental setups allowing a more robust control over various parameters. Notably, Van den Berg et al. (1959) used a setup with extensive measurement of the airflow rate delivered to the larynx and the pressure just below the glottis (subglottal pressure), while elongating the VFs (by applying longitudinal force on the thyroid cartilage and rotate it with respect to the cricoid cartilage – recall Figure 3) and applying lateral forces on the arytenoids to adduct the VFs (to simulate the action of the adductor laryngeal muscles – recall Figure 4).

The experimental setup used by Van den Berg et al. (1959) formed the basis of excised larynx experimental setups commonly used in later studies (Döllinger et al., 2011; Garcia Maxime et al., 2018). Its principle is rather straightforward: airflow is generated by a high-pressure container, then goes through another large container where it is heated and saturated in humidity, to finally be delivered to the larynx (see Figure 20). Van den Berg also included a way to measure the subglottal pressure and the temperature by connecting the subglottal tract to a U-tube and to a thermometer, as well as a system of weights and pulleys to control and monitor the forces applied for the elongation of the VF (longitudinal force) and for the arytenoid adduction (lateral force), as shown on Figure 20. The setup finally includes a microphone to register the radiated pressure, and a stroboscope to allow the visual observation of the VF vibrations.

Van den Berg further improved this experimental setup by adding sutures that mimic the action of some intrinsic laryngeal muscles (recall Figure 4). Those sutures are usually attached directly to the arytenoids, where the muscles normally attach. Such suture attachments are documented with remarkable details in the instructional film from Van den Berg et al. (1960), where the action of all three intrinsic laryngeal muscles described in Figure 4 (LCA, PCA and IA) is demonstrated, both independently and simultaneously. The action of the CT muscle is also mimicked by pulling a suture thread attached to the anterior notch of the thyroid cartilage (as depicted in Figure 3), which demonstrates an increase of f_o. Technological advances in electronics allowed recent experimental setups to include electric servo-motors coupled with force sensors, to more precisely control and monitor the lateral (adduction/abduction) and longitudinal (VF elongation) forces (Legou et al., 2015; Lagier et al., 2017; Lagier et al., 2020). This setup also controls all servo-motors centrally through an Arduino board connected to a computer, allowing more precision and reproducibility on the force controls, as well as their automation.

One common limitation of excised larynx experiments is the difficulty to mimic the action of the TA muscle. Montgomery implants (solid wedges inserted through the wings of the thyroid cartilage) have been used to adduct the membranous part of the VFs (Lagier et al., 2017), but the tensing and bulging effect of the TA muscle contraction cannot be simulated without innervation and electrical stimulation of the muscle. This was successfully performed by Berke et al. (2013), who demonstrated and documented the visible adduction motion resulting from the electrical stimulation of the TA muscle in a perfused human excised larynx. However, the procedure requires considerable preparation and knowledge regarding blood perfusion, making this kind of experiment difficult to set up.

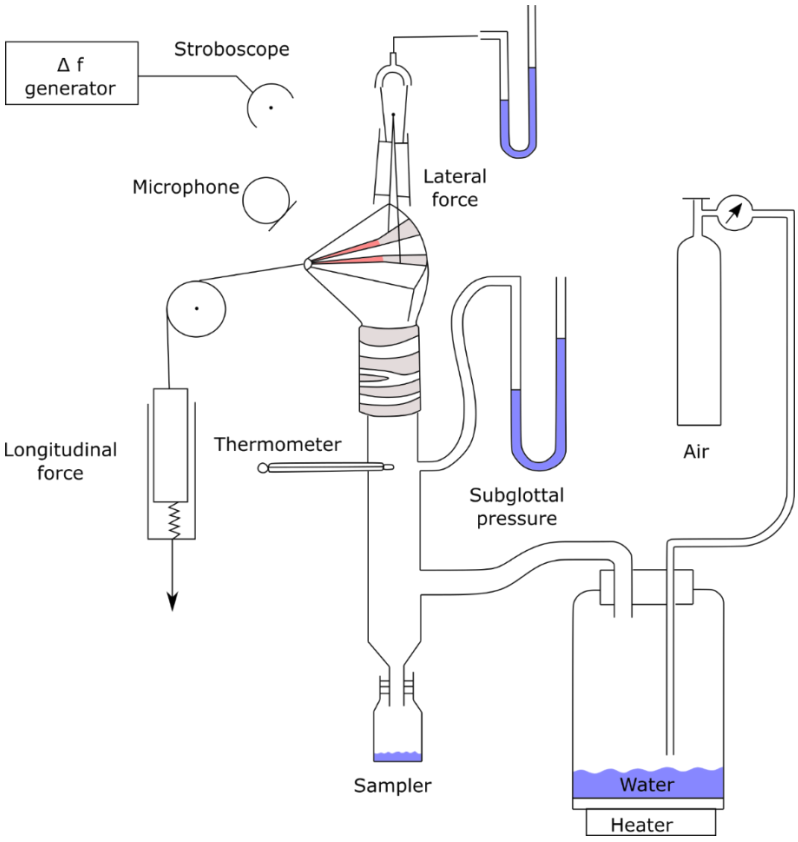


Figure 20. The setup used by Van den Berg for excised larynx experiments. Based on Van den Berg et al. (1959), Fig. 1, modified by the author.

Some substantial advances in our understanding of the voice production mechanisms were made through the use of hemilarynges, which are excised larynges where one VF is surgically removed, allowing the observation of the VF vibrations from the top and from the side (Jiang et al., 1993; Herbst et al., 2017). Notably, the relationship between the vocal fold contact area and the EGG signal was explored using hemilarynges (Scherer et al., 1988; Hampala et al., 2016). In addition, the modes of vibration of the VFs (recall section 2.3.4) were investigated by attaching micro-sutures to the VF medial surface and analyzing the motion of those micro-sutures through HSV (Berry et al., 2001; Döllinger et al., 2006; Veltrup et al., 2021).

Even though human excised larynges offer the best model for investigating human voice production and physiology, they still pose ethical concerns as they are obtained from human remains. Another limitation of excised human larynges is their age, as they are most often obtained from older individuals who died of natural cause. As age progresses, the cartilages ossify and become harder, which can make it more difficult for the preparation. To overcome those limitations, researchers have been using various mammalian larynges as replacements, such as larynges from dogs, cows, sheep, pigs, or deer. Those species seem to offer relatively similar mechanical properties of the larynx, although differences exist and should be taken into account (Alipour et al., 2013; Herbst et al., 2020). Other species have been studied, but the description of those studies is beyond the scope of this thesis. Readers can refer to the review from Garcia et al. (2018) for a more comprehensive review on excised larynx experiments.

2.5.2. Larynx storage and preparation

Larynges must ideally be excised and stored within 24 hours post-mortem (Titze, 2006a). Once larynges are harvested, they are usually flash-frozen in liquid nitrogen, which can prevent cracks, and then stored in freezers at low temperature (from -120°C to -18°C). If quick frozen in liquid nitrogen and stored properly in vacuumed bags, excised larynges can be kept for up to one month and used for experiments without substantial deterioration of the tissues altering the voice production (Chan et al., 2003). The effects of a more prolonged storage on the biomechanical properties of the VFs are unknown, however. When used for experiments, a larynx should be slowly thawed by being placed in a fridge at a temperature around 0°C overnight, and then placed in a warm water bath with saline solution at around 30°C (Baer, 1975; Durham et al., 1987; Titze, 2006a), in order to avoid any cracks or damage to the tissues caused by quick temperature variations.

Once the larynx is fully thawed, the usual preparation consists in removing superficial tissues and cartilages that are not needed for the experiment, or that can impede the view of the VFs. Most commonly, the tissues above the VFs, such as the epiglottis, the hyoid bone, and sometimes the ventricular folds, are removed. If present, unnecessary surrounding tissues, such as the esophagus, can also be removed for facilitating the access to the laryngeal muscles and cartilages. The trachea should be kept or shortened according to the needs of the experimental setup. If needed, one VF can be removed to transform the larynx into a hemilarynx.

Once the larynx is prepared according to the needs of the experiment, it is attached to the experimental setup and the VFs should be adducted to create a sufficient airflow resistance, which will put them into oscillation. As mentioned in section 2.5.1, the arytenoid adduction can be performed by mimicking the action of some of the intrinsic laryngeal muscles (LCA and IA, for example), which is achieved through threads attached to the arytenoids and going through the tissue in the direction of the muscle. Figure 21 shows photos of an excised red deer larynx used by the author HL, with threads attached to the arytenoids to mimic the action of the LCA muscle. In this case, the threads go through the tissues in a forward and downward direction, to come out through the crico-thyroid membrane. They are also attached to a small bead to prevent the threads from losing their attachment, and to push the arytenoids towards the direction of the threads when pulled, which has the effect of rotating the arytenoids and adducting the VFs.

Another possibility for the VFs adduction is to use a system of pins or prongs that push against the arytenoids to create the adduction movement. The example of such a system is described by Titze (2006a). Figure 22 depicts the system that was used to adduct the VFs during the excised larynx experiments in papers **I** and **II** of this thesis: threaded metal prongs were screwed into a plate attached to two metal rods. Figure 22a shows the resting position of the arytenoids without the prongs exercising any force on them. Figure 22b shows the arytenoid adduction when the prongs are pressed against them: the adjustment of the prongs causes the arytenoids to come closer together and to slightly rotate.

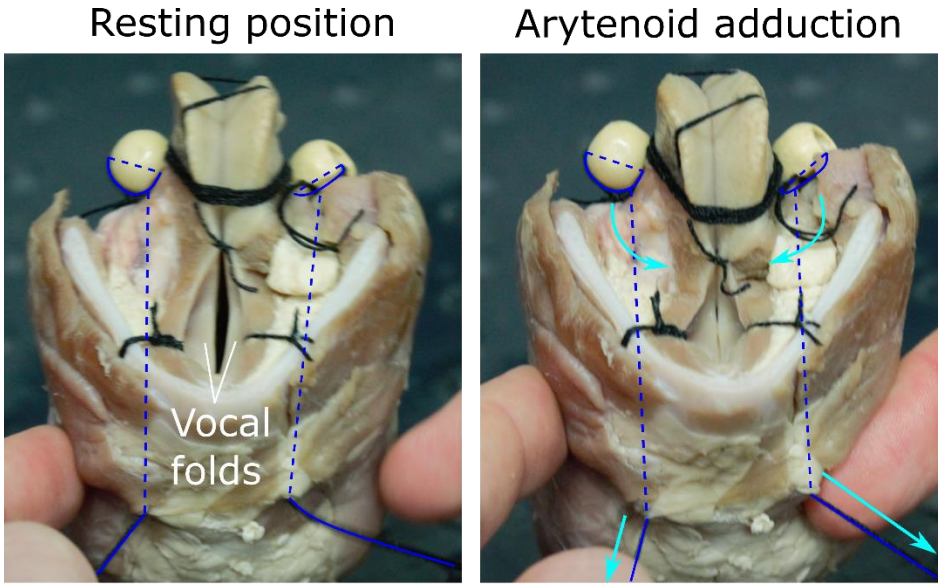


Figure 21. Excised red deer larynx with threads (in blue) attached to the arytenoids. The left panel shows the resting position of the larynx, the right panel shows the adduction process by pulling on the threads.

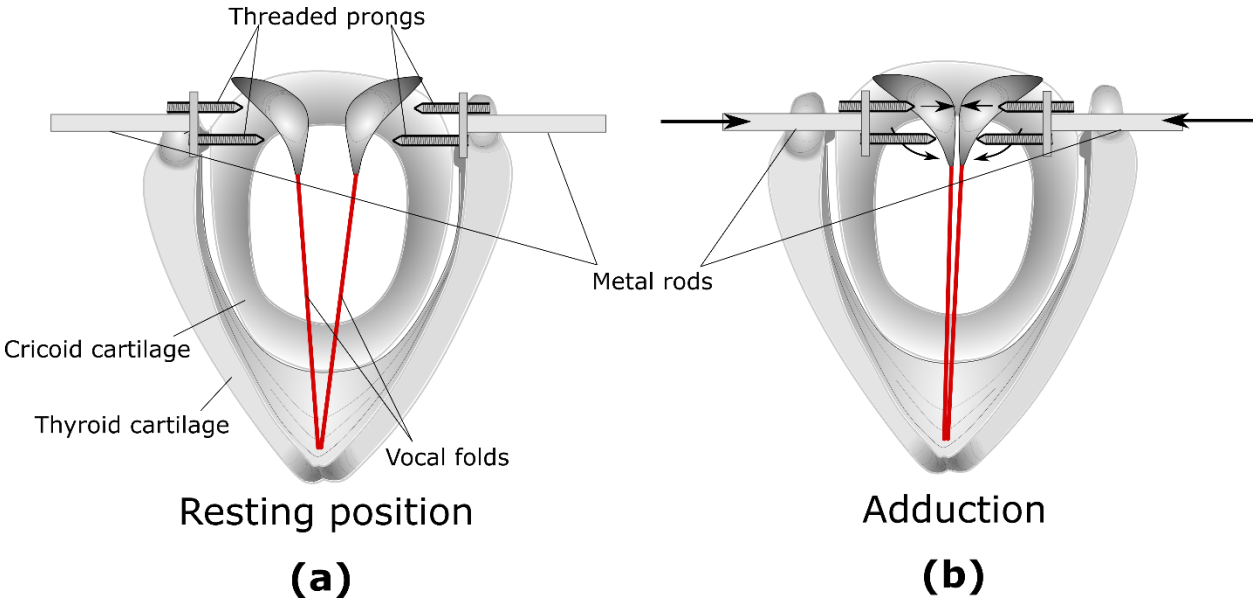


Figure 22. Description of the VF adduction using prongs. (a) resting position showing the prongs screwed into a metal rod; (b) the metal rods are pushed medially, which pushes the prongs against the arytenoids and causes them to come closer and to rotate.

2.5.3. Investigation of nonlinear dynamic phenomena in excised larynges

As previously mentioned, excised larynges offer the possibility to control, in a reproducible way, several parameters which are difficult to control in vivo. Such parameters include the airflow rate, and the VF adduction and elongation. The subglottal pressure can be indirectly controlled through the airflow rate and the flow resistance caused by the VF adduction. The degree of left-right VF asymmetry with regards to the VF adduction and elongation can also be controlled. Furthermore, the characteristics of the subglottal and supraglottal tracts, and the way they connect to the larynx, can be controlled to some

extent. Therefore, excised larynges can be used to investigate nonlinear phenomena related to the inherent laryngeal characteristics and to the coupling of the VF vibrations with the supraglottal or subglottal tract.

Particularly, frequency jumps (recall sections 2.3.1 and 2.3.4) have been investigated during excised larynx experiments, by elongating the VFs (Van den Berg et al., 1960; Švec et al., 1999; Horáček et al., 2004; Tokuda et al., 2007, 2008) or by inducing a left-right asymmetry in the VF adduction and elongation (Berry, Herzel, et al., 1996). The instructional film by Van den Berg et al. (1960), particularly, clearly demonstrates frequency jumps (associated with a register transition), when the VFs are smoothly elongated by a thread pulling and rotating the thyroid cartilage. This suggests that such jumps can also occur without any acoustic interactions with the supraglottal tract. Furthermore, since the TA muscle could not be stimulated in those studies, this implies that the sudden release of the TA activity (as hypothesized by Titze (2000b)) is not necessary for frequency jumps to occur. Finally, the study from Berry, Herzel, et al. (1996) demonstrates that sufficient left-right asymmetry in the VF adduction and elongation can disrupt eigenmode entrainment (recall section 2.3.4), which stimulates eigenmode desynchronization and vocal instabilities.

2.6.Synthesis of kymograms through numerical modelling

Numerical models of the VFs can offer more control on the VF geometric and vibratory properties than physical models, and as such their use can be relevant in a research context. Specifically, numerical models can be used as a visualization tool to simulate images of the VF vibrations, such as kymograms or high-speed videos of the vibrating glottis. The development of videokymography (VKG – recall section 2.2.1.5), in the late 1990s and early 2000s, particularly stimulated the need to use numerical modelling for the generation of synthetic kymograms, to improve the interpretation of kymographic images and relate them to different types of phonation.

Notably, Horáček et al. (2009) were among the first to generate synthetic kymograms, using an aeroelastic model of the VF vibrations. This model approximates the VFs in two dimensions (2D) by a rigid body defining their shape, connected to a two-degree-of-freedom (translational and rotational) system of masses, springs and dampers (Horáček et al., 2002). Here, the VFs are driven by aerodynamic forces generated by the airflow, and kymograms were generated by placing a virtual camera above the VFs and using an illumination method to measure the quantity of light reflected by the VF surface, based on the local surface slope. Three examples of synthetic kymograms generated by this model are displayed on Figure 23. This model is able to generate simple kymograms where both the upper and lower margins of the VFs are visible (Figure 23a), however since the VFs are approximated by a rigid body, the mucosal wave extent (the propagation of the mucosal wave across the top surface of the VFs) is not visible and the kymograms are visually quite distinct from those obtained from real VFs in vivo. Nevertheless, this model allows to investigate the appearance of kymograms when varying the subglottal pressure (Figure 23b) and the frequency of oscillation f_o (Figure 23c).

Later, Švancara et al. (2014) performed a similar study with a 2D finite element model, where the VF shape is based after the M5 model defined by Scherer et al. (2001). The VFs are discretized using a finite element mesh, and set to vibration by fluid-structure interaction with the airflow. Kymograms are generated using the same illumination method used in Horáček et al. (2009). As the VFs are deformable in this model, the mucosal wave extent is visible on the generated synthetic kymograms, an example of which is displayed

on Figure 24. This kymogram has a more realistic appearance than the kymogram generated by Horáček et al. (2009) but the vibratory features are still quite distinct from those visible in the kymograms of real VFs.

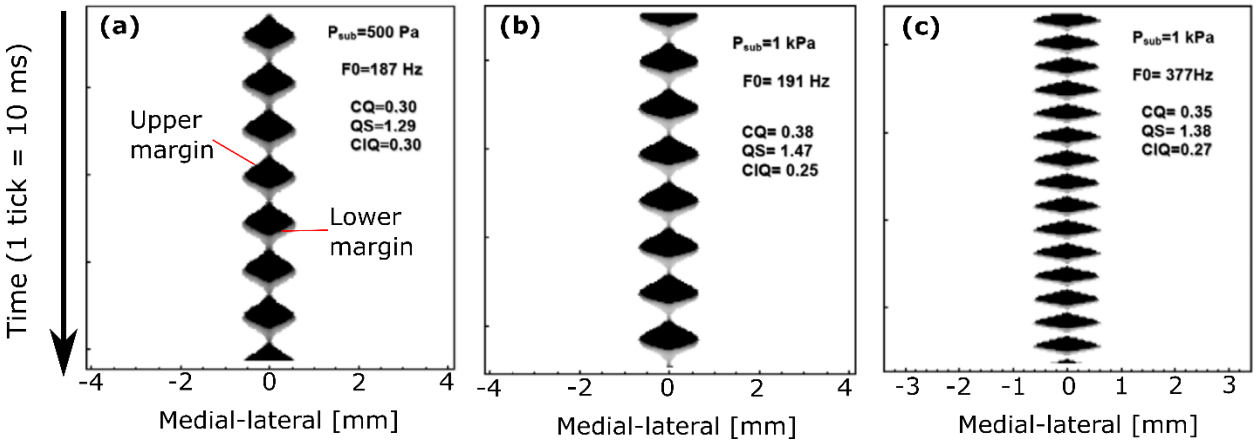


Figure 23. Example kymograms generated by an aeroelastic model of the VFs. The contributions to the kymogram from the upper and lower margins are indicated with red lines. The values of the closed quotient (CQ), the speed quotient (QS), and the closing quotient (ClQ), are also indicated. (a) Kymogram generated with a subglottal pressure of 500 Pa, with the frequency of oscillation of 187 Hz; (b) The subglottal pressure is increased to 1 kPa, which has the effect of increasing the vibration amplitude and f_o; (c) The subglottal pressure is kept at 1 kPa, but f_o is increased to 377 Hz by increasing the VF stiffness. Modified by the author from Horáček et al. (2009).

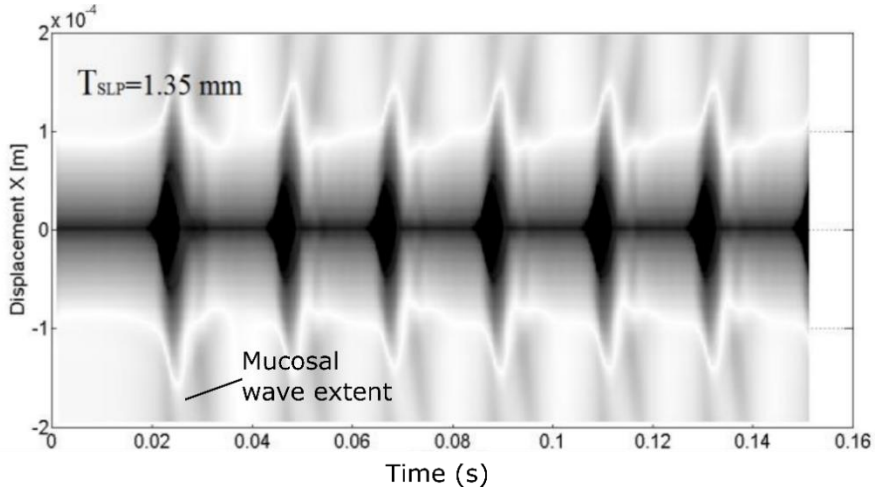


Figure 24. Kymogram generated from a finite element model of the VF vibrations. The thickness of the lamina propria T_{SLP} is indicated here. Modified by the author from Švancara et al. (2014).

These two models involve solving complex equations of fluid-structure interactions, which are computationally expensive, especially for finite element models containing a high number of points for calculation. To address this potential limitation, kinematic VF models have been designed for simulating kymograms. Here, the VFs are set in vibration by applying kinematic rules instead of driving them with aerodynamic forces. Besides of computational simplicity, this also has the benefit of providing more direct control on the pattern of vibration. Notably, Schoentgen et al. (2015) used the phase-delayed overlapping sinusoids (PDOS) model, which applies a medial-lateral sinusoidal motion to the VF surface, where the glottal exit (upper margin) is phase delayed with respect to the glottal entry (lower margin) (Titze, 2006b).

Schoentgen et al. (2015) demonstrated that the model is capable of generating synthetic kymograms exhibiting left-right asymmetry, which have a more realistic appearance since the VFs vibrate with some degree of asymmetry in vivo. However, this model cannot display the mucosal wave extent, as the motion of the VFs is only horizontal.

The recent mucosal wave model developed by Kumar et al. (2019) addresses this issue by applying circular (lateral-medial and vertical) motion to the VFs. This model discretizes the VF surface into equidistant points following the M5 shape, which can be fully defined by a few geometric parameters. Each point is then set to motion following a circular trajectory, with a constant phase delay between all successive points. The definition of the model using only a few geometric (glottal width, convergence angle of the medial surface) and kinematic (amplitudes and frequencies of vibration, phase difference) parameters allows for a high flexibility in the appearance of the generated kymograms, which can simulate various types of phonations and offer the closest resemblance to kymograms of real VFs to date, also with regards to mucosal waves travelling over the top VF surface. This is demonstrated in Figure 25, which displays kymograms corresponding to a soft (Figure 25a), pressed (Figure 25b), breathy (Figure 25c) and head (Figure 25d) voice. This model uses a more complex illumination method which also considers the vertical distance from the camera.

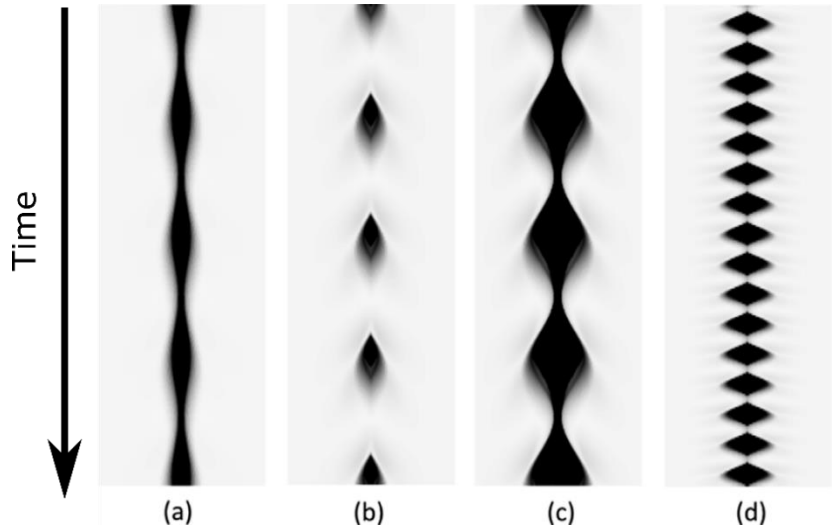


Figure 25. Synthetic kymograms generated by the mucosal wave model, showing different simulated types of phonation: (a) soft, (b) pressed, (c) breathy, and (d) head register. Modified by the author from Kumar et al. (2019).

3. ORIGINAL WORK BY THE AUTHOR

3.1. Aims of this thesis

The aims of this thesis are:

- to test and document a newly developed anechoic subglottal tract, and to investigate the behavior of the VF vibrations using excised larynges in fully anechoic conditions and compare it with subglottally resonant conditions. In particular, the aim is to document the differences in the subglottal and radiated pressure waveforms. This is addressed in paper **I**.
- to investigate the differences in the occurrences of frequency jumps in human excised larynges, between anechoic and subglottally resonant conditions. This is addressed in paper **II**.
- to investigate the laryngeal differences in vivo between chest and head registers throughout a wide singing range, and to find relevant glottal parameters that can discriminate between the two registers throughout this entire range. This is addressed in paper **III**.
- to help to extend a previously developed two-dimensional numerical kinematic model of the VF vibrations to three dimensions, so that it can generate not only kymograms but also synthetic videos of the entire glottis, which appear similar to laryngoscopic videos obtained in vivo during laryngeal HSV. Particularly, the aim is to allow the visualization of realistic vibrations, containing anterior-posterior and left-right differences. This is addressed in paper **IV**.

This section gives a summary of the work by the author on each paper in this thesis, and gives an overview of their most significant results. Readers are referred to the papers themselves for more detailed descriptions.

*3.2. Paper **I**: Development of an anechoic subglottal tract, and comparison of subglottal pressure waveforms in anechoic and subglottally resonant conditions*

3.2.1. Objectives

The glottal flow and the VF vibrations have been shown to be influenced by the acoustic resonances of both the supraglottal (Titze, 2008; Titze et al., 2008; Tokuda et al., 2010; Zañartu et al., 2011; Wade et al., 2017) and subglottal (Zhang et al., 2006) tracts. The study the inherent vibratory properties of the VFs and of the glottal voice source, it is necessary to remove all acoustic interactions with the supraglottal or subglottal tract. This study proposes to address this need by using excised larynges (which are free of a supraglottal tract) and a newly developed anechoic subglottal tract, which is described here. The influence of the subglottal resonances is further explored by comparing the subglottal pressures and properties of the VF vibrations in anechoic and resonant conditions. Specifically, the subglottal pressure waveform is investigated. In addition, the influence of the subglottal resonances on the vibratory pattern of the VFs (referred to as level 2 interactions by Titze (2008)) is assessed by measuring the EGG signal and the fundamental frequency of oscillation f_o.

3.2.2. Design of the anechoic subglottal tract

The anechoic subglottal tract was constructed in the Voice Research Laboratory at the Palacký University in Olomouc by the former doctoral student V. Hampala under the supervision of J. G. Švec. The tract consists of a 3.30 m plastic tube terminated by a wedge of polyurethane foam to absorb the sound waves

propagating through it, which prevents their reflection and therefore suppresses the acoustic resonances. Detailed description, analysis of the properties and behavior of the tract was carried on by the author.

To verify that the anechoic subglottal tract was suppressing the acoustic resonances, its acoustic response was first measured and compared with the acoustic response of the resonant subglottal tract previously developed by Hampala et al. (2013). It was found that the anechoic subglottal tract is successfully suppressing the acoustic resonances, as shown on Figure 26: the resonant subglottal tract exhibits visible resonances and antiresonances, whereas the anechoic subglottal tract is practically free of these.

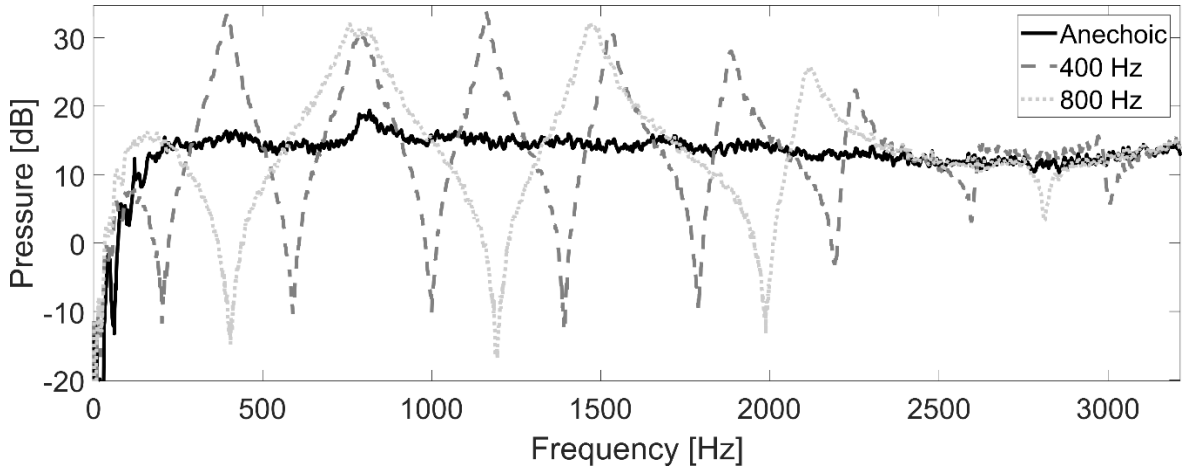


Figure 26. Frequency responses of the anechoic subglottal tract and of the resonant subglottal tract with the piston set to two different positions corresponding to fR1 = 400 and 800 Hz.

3.2.3. Excised larynx experiments

Red deer excised larynges were then explored during two experiments, using the experimental setup depicted on Figure 27. A pressure sensor was inserted through the dorsal ridge of the cricoid cartilage to measure the mean and oscillating subglottal pressure, and a microphone was placed approximately 10 cm away from the glottis to register the radiated sound. EGG electrodes were also placed on each side of the thyroid cartilage, at the level of the VFs, and connected to the Glottal Enterprises EG2-PC device to monitor the relative VF contact area. For both experiments, the value of f_o was measured using the SWIPE' algorithm developed by Camacho et al. (2008).

The first experiment investigated steady phonations (one larynx), in both anechoic and resonant conditions. In resonant conditions, the subglottal tract was set to six different resonance frequencies: f_{R1} = 330, 400, 500, 600, 700 and 800 Hz. The flow was increased to a constant mean value of about 400 mL.s^{-1}, until the mean subglottal pressure reached a saturation value, and then the flow was decreased back to 0 mL.s^{-1}. This experiment investigated the differences in the subglottal pressure waveforms between anechoic and subglottally resonant conditions.

The second experiment investigated flow sweeps (three larynges), again in both anechoic and resonant conditions. For this experiment, in the subglottally resonant conditions the resonance frequency was only set to f_{R1} = 500 Hz, to approximate the value of the subglottal resonance frequency in vivo (see section 2.1.1.3). The flow was increased slowly to the maximum value (550 mL.s^{-1}) and then slowly decreased back to 0 mL.s^{-1}. In this experiment, the phonation threshold pressure, and the SPL of the subglottal and

radiated sounds, were also measured and compared between anechoic and subglottally resonant conditions – for details see the paper **I**.

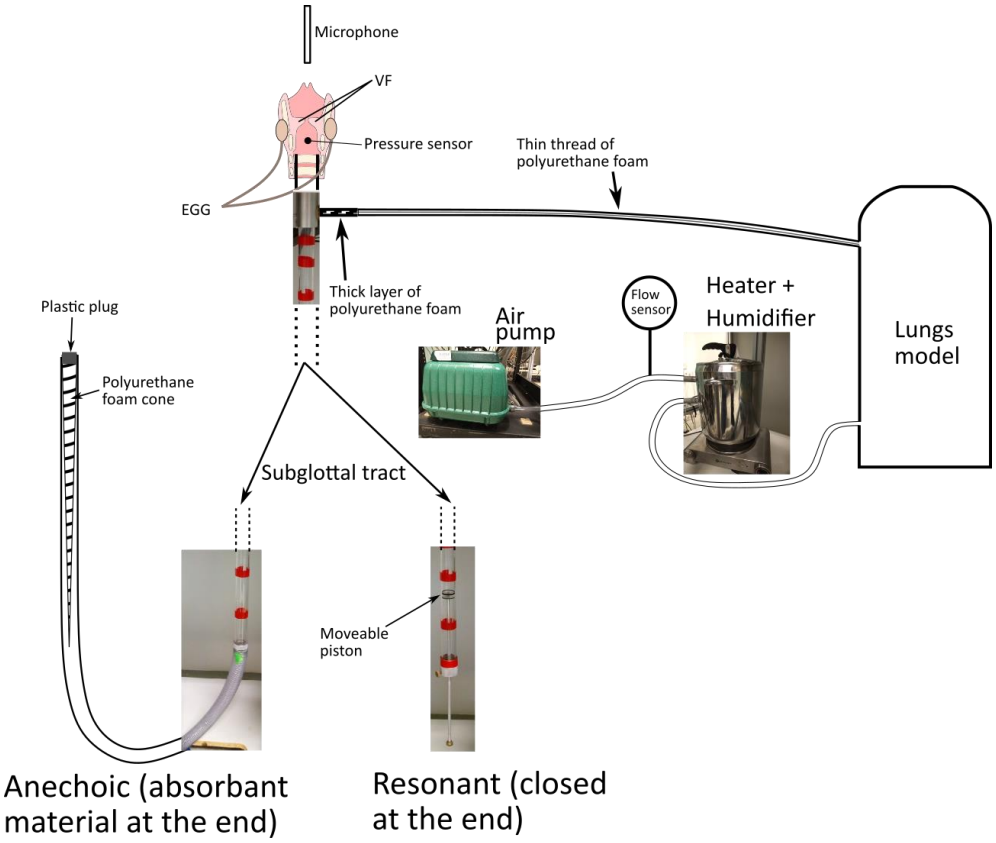


Figure 27. Excised larynx experimental setup.

Remarkable differences were visible in the subglottal pressure waveforms between anechoic and subglottally resonant conditions, shown in Figure 28. As expected, in the anechoic conditions, the subglottal pressure waveform was found to be similar to the inverted theoretical glottal flow (Sundberg, 2018), as shown on Figure 28a: it was decreasing during the opening phase, increasing during the closing phase, and was rather flat with little fluctuations during the closed phase (recall section 2.2.3.1). On the other hand, in the resonant conditions, the subglottal pressure waveforms showed strong fluctuations during the closed phase, and the frequency of those fluctuations increased as the subglottal resonance frequency increased, as demonstrated for f_{R1} = 330 Hz (Figure 28b) and for f_{R1} = 500 Hz (Figure 28c). In addition, the EGG waveforms also exhibited, especially for f_{R1} = 330 Hz (Figure 28b), strong fluctuations in the closed phase, which appear to be synchronized with the fluctuations in the subglottal pressure waveform.

In both experiments, the values of f_o were consistently lower in the subglottally resonant conditions than in the anechoic conditions. The values measured during the second experiment are plotted against the mean subglottal pressure, for larynx #1 (only the values for the first and last repetitions are shown for brevity), in Figure 29.

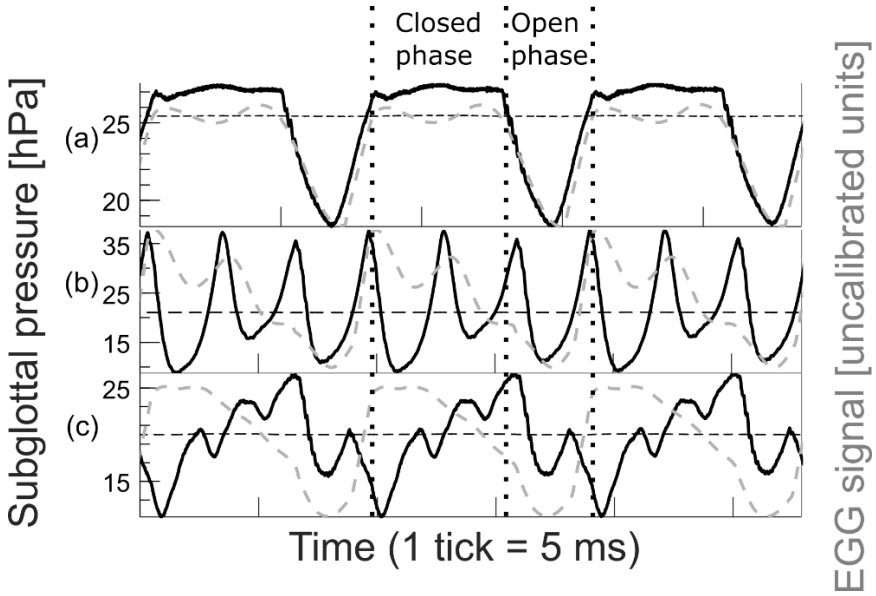


Figure 28. Subglottal pressure (solid black lines) waveforms registered with an anechoic subglottal tract and with resonant subglottal tracts set to different resonance conditions: (a) anechoic subglottal tract, (b) f_{R1} = 330 Hz, (c) f_{R1} = 500 Hz. The horizontal dashed black lines indicate the mean subglottal pressure in each case. The waveforms are individually scaled in time (3 cycles are shown). The instants of closure and opening were approximately synchronized using the corresponding EGG signal waveforms (dashed gray lines). Finally, the closed and open phases are indicated by vertical dotted black lines.

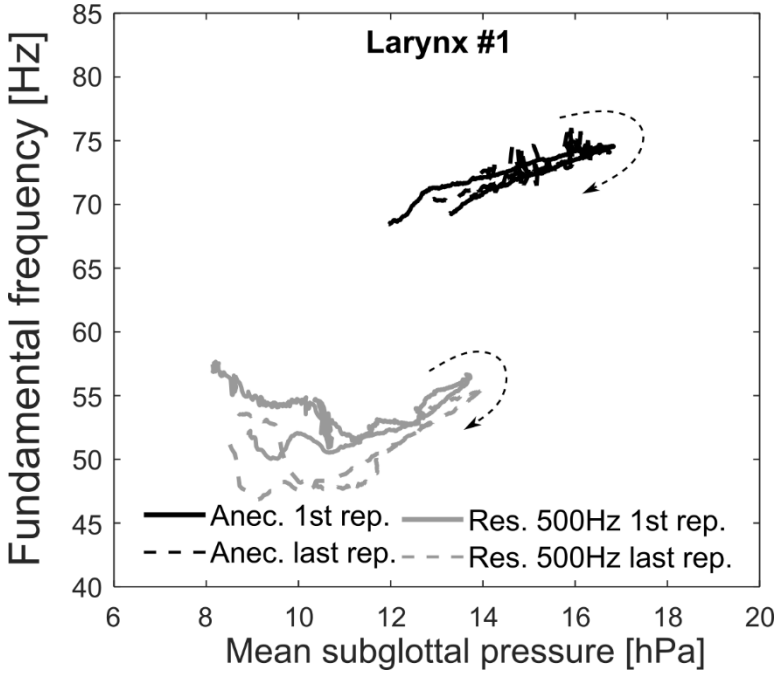


Figure 29. Measurement of the fundamental frequency of oscillation f_o made during the flow sweep experiments, for larynx #1. The dashed black arrows indicate the evolution of the measured values with time. Notice the different f_o values in anechoic and resonant conditions.

3.2.4. Discussion

The resemblance of the subglottal pressure waveform to the theoretical glottal flow waveform in anechoic conditions (recall Figure 15a) goes in line with the theoretical relationship between the flow and the

pressure (described in the Appendix of paper **I**) in the absence of acoustic resonances. The flat subglottal pressure waveform during the closed phase of glottis provides a direct visual evidence of the absence of subglottal resonances and proper functioning of the anechoic subglottal tract. On the other hand, the subglottal pressure fluctuations observed during the closed phase in the resonant cases indicate the presence of subglottal resonances, as reported previously by measurements in vivo (Miller et al., 1985; Schutte et al., 1988; Švec et al., 2021).

The fluctuations in the EGG signal and the lower values of f_o measured in the resonant conditions provide direct evidence of level 2 interactions due to the presence of subglottal resonances. This corroborates the findings from Zhang et al. (2006), who also reported a strong influence of subglottal resonances on the VF oscillatory frequency f_o.

3.2.5. Conclusion

The newly developed anechoic subglottal tract successfully suppresses its acoustic resonances, removing their influence on the VF vibrations. In anechoic conditions, the subglottal pressure waveform resembles the inverted theoretical source flow waveform (recall Figure 15a), which allows to study the source signal without the need for inverse filtering. The presence of subglottal resonances consistently lowers the fundamental frequency of oscillation of the VFs and generates strong fluctuations in the EGG signal during the closed phase, which provides evidence of level 2 interactions with the subglottal pressure oscillations. The developed anechoic subglottal tract can be used to investigate the inherent vibratory properties of the voice source and the impact of structure-acoustic interactions with subglottal resonances on those properties.

*3.3.Paper **II**: Frequency jumps in anechoic conditions*

3.3.1.Objectives

Frequency jumps are a common vocal phenomenon which occurs when young prepubescent individuals go through puberty, or when singers switch from one register to another. Frequency jumps have been explained through several hypotheses: sudden release of the TA muscle (recall section 2.3.1), interactions with the supraglottal resonances (recall section 2.3.2) and with the subglottal resonances (Titze, 2000b), or nonlinear phenomenon caused by VF eigenmode desynchronization (recall section 2.3.4). However, the exact mechanism responsible for frequency jumps is still unknown. Past studies have demonstrated frequency jump events in excised larynges without supraglottal tract, which indicates that supraglottal resonances are not needed. However, as mentioned in section 3.2, subglottal resonances can have an influence on the VF vibrations, and as such could be responsible for the occurrence of frequency jumps in excised larynges. Consequently, the question of whether frequency jumps can also occur in fully anechoic conditions remains unanswered. This study proposes to answer this question by exploring frequency jumps in human excised larynges in anechoic conditions, using the anechoic subglottal tract described in paper **I** (see section 3.2 for a brief summary) and no supraglottal tract. The influence of the subglottal resonances on frequency jumps is also investigated by using a resonant subglottal tract.

3.3.2.Methods

Five human excised larynges are used with the experimental setup described in paper **I** (see Figure 27). The frequency of oscillation f_o is increased and decreased by smoothly elongating and shortening the VFs, which is done by manually pulling the anterior notch of the thyroid cartilage (see Figure 30).

Elongation/shortening sweeps are performed in anechoic condition and in subglottally resonant conditions with the first subglottal resonance frequency set to f_{R1} = 500 Hz (i.e. similarly as observed in the *in vivo* conditions), and repeated to assess reproducibility

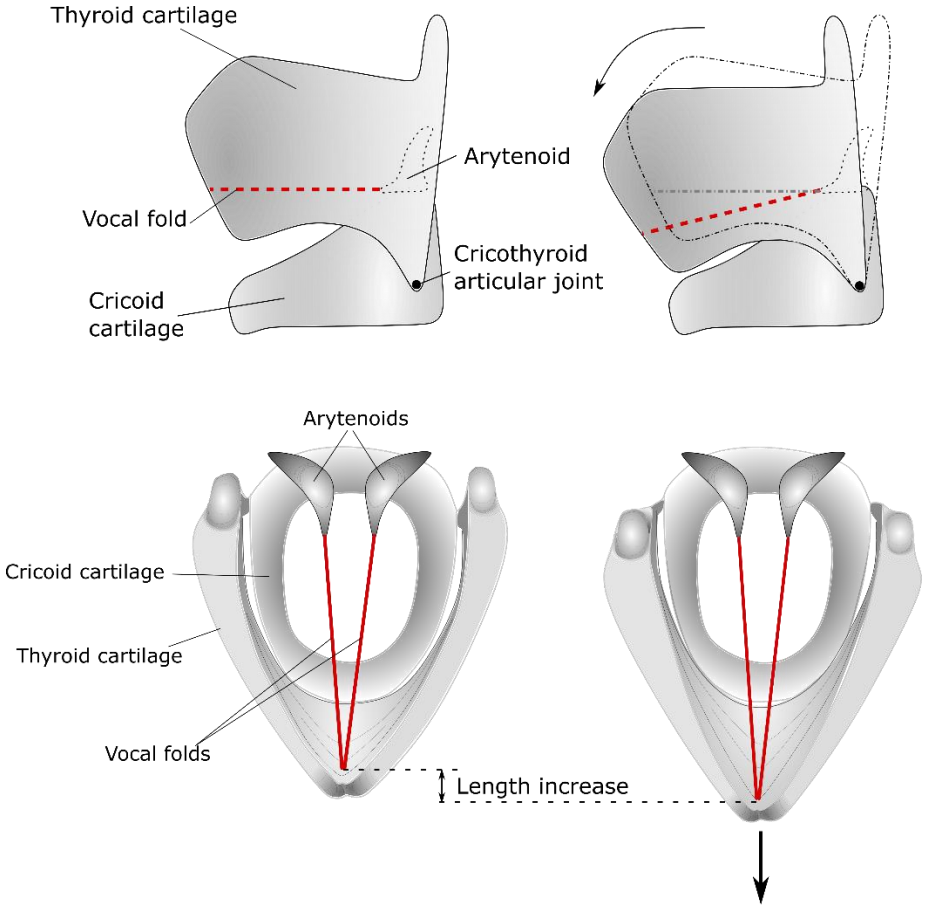


Figure 30. Description of the VF elongation process by pulling on the anterior notch of the thyroid cartilage.

3.3.3.Results

When smoothly elongating the VFs, consistent jumps were observed in both anechoic and subglottally resonant conditions, for two larynges. Figure 31 shows the values of f_o right before and after each jump, for larynx #1 and larynx #2, in both anechoic and resonant conditions. The presence of subglottal resonances did not stimulate more numerous jumps than in anechoic conditions, however the

resonant subglottal tract slightly altered the starting and terminating frequencies of the jumps, especially during the upward jumps (Figure 31a,c).

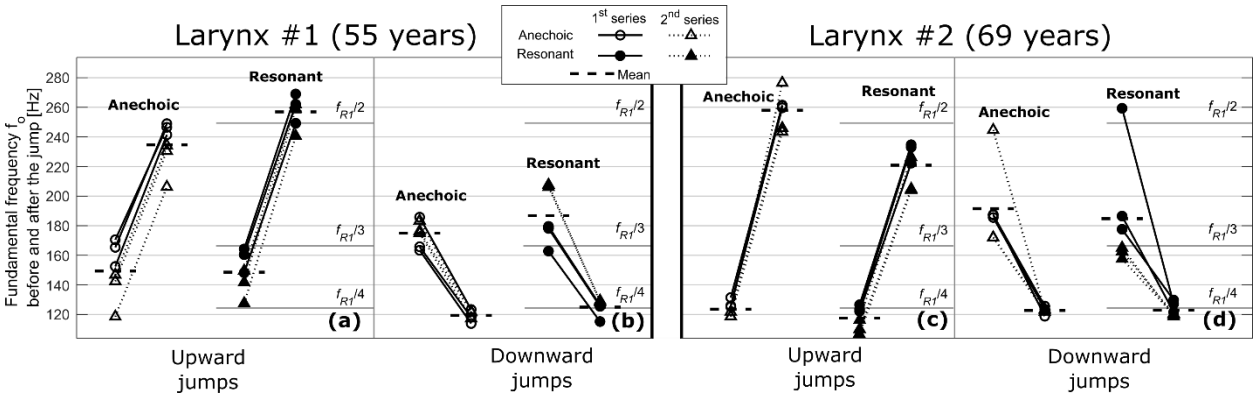


Figure 31. Fundamental frequency (f_o) before and after each frequency jump for larynx #1 and #2. The circles correspond to the first, and the triangles to the second series of frequency sweeps. For the resonant cases, we also indicate the integer divisions of the first subglottal resonance frequency (f_{R1} ≈ 500 Hz), $f_{R1}/2$, $f_{R1}/3$, and $f_{R1}/4$. (a) larynx #1, upward jumps; (b) larynx #1, downward jumps; (c) larynx #2, upward jumps; (d) larynx #2, downward jumps.

3.3.4.Discussion and conclusion

This study presents the first clear evidence that frequency jumps occur in excised human larynges also in fully anechoic conditions, suggesting that acoustic interactions with the supraglottal and subglottal tracts are not necessary for those jumps to occur. Furthermore, the presence of acoustic resonances in the subglottal tract did not stimulate more numerous jumps compared to anechoic conditions, which confirms that those jumps were primarily caused by inherent VF properties rather than by acoustic interactions with the subglottal tract. Nevertheless, the presence of acoustic resonances in the subglottal tract had an impact on the terminating frequencies of the upward jumps, which suggests that the subglottal resonances also have an influence on the frequency jumps, although this influence appears secondary. As such, this study provides important insights on the underlying mechanisms behind frequency jumps.

*3.4. Paper **III**: Laryngeal adjustment differences between chest and head registers*

3.4.1.Objectives

As mentioned in section 2.4, the exact underlying mechanisms behind the production of voice registers are not fully understood. Frequency jumps described in paper **II** (see section 3.3 for a brief summary) are often associated with a transition between two registers, even though singers are able to operate a smooth pitch glide without any perceivable transition event, by precisely controlling the adjustments of the laryngeal muscles (recall section 2.3.1). These adjustments, and particularly the activity of the TA muscle, have been hypothesized to be essential in producing different registers, specifically chest and head registers. Previous studies have investigated the differences in laryngeal adjustments between chest and head registers (recall section 2.4.2), but only over a limited pitch range where both registers can be produced. However, the pitch limitation of chest register to low pitches and of head register to high pitches could be challenged. There is even evidence that both registers can be produced over a wider frequency range, for example as demonstrated by the Australian singer Mal Webb, who performs register transitions while singing the same note, in his YouTube video “Sideways yodeling” (Webb, 2008).

Overall, a way to reliably recognize chest and head registers, and differentiate one from the other independently from pitch, is still lacking, particularly regarding the laryngeal adjustments employed by trained singers to produce each register. Inspired by the Mal Webb idea, the American singer Lisa Popeil came to the Voice Research Lab at the Palacký University in Olomouc with the idea of separating registers from pitch: she claimed that she can produce both the chest and head registers throughout her entire singing range. This study investigates this claim in detail, with the aim of finding relevant laryngeal and glottal parameters that can allow the discrimination between chest and head registers, regardless of pitch. The investigation is based on several hypotheses based on previous studies reporting more TA muscle activity and a richer harmonic structure of the sound in chest register. Those hypotheses are described in detail in paper **III** in this thesis but can be summarized as follows: in chest register, the VFs will exhibit a longer and more pronounced contact, larger phase differences between their lower and upper margins, a higher maximum closing speed, and there will be more high-frequency energy in the radiated sound spectrum.

In this study, the terms “chest-like” and “head-like” are used to refer to the singer’s intended registers, to avoid confusion with the pitch-associated “chest” and “head”. The terms “chest” and “head” are only used here to refer to the registers described previously in the literature.

3.4.2.Methods

The professional singer (co-author LP) performed short sustained phonations alternating chest-like and head-like registers with a short pause, on every pitch from C3 (131 Hz) to C6 (1047 Hz), except for Ab3 (208 Hz) which was omitted by mistake. The VF vibrations are captured by a high-speed camera connected to a 90° endoscope, at the rate of 7200 frames per second (FPS) for the pitches from C3 to Db5 (554 Hz), and 13600 FPS for the pitches from D5 (587 Hz) to C6. The VF contact and the radiated sound are also registered through EGG electrodes placed on the larynx and through a microphone placed on top of the camera at the distance of 22 cm from the mouth. The microphone sound is used for blind listening tests performed by the three authors of the study (“insiders” group) and by six other participants who were not familiar with the study design (“outsiders” group). The listening tests were performed to determine whether the intended registers could be perceptually distinguished. The experimental setup is shown in Figure 32. The following parameters are calculated from the phonovibrograms (PVG), the glottal area waveform (GAW) and the kymographic contour of the vibrating VFs:

- Closed quotient from the PVG (CQ_{PVG})
- Speed quotient from the kymographic contour (SQ_{kym})
- Closing quotient from the kymographic contour (CgQ_{kym})
- Vertical phase differences between the lower and upper margins of the VFs from the kymographic contour (VPD_{kym})
- Normalized amplitude quotient from the GAW (NAQ_{GA}).

In addition, the following parameters are calculated: the amplitude of the EGG signal, the sound pressure level (SPL), the spectral balance (SB), and the difference in the levels of the first and second harmonics L_{H1}-L_{H2}– see paper **III** for details.

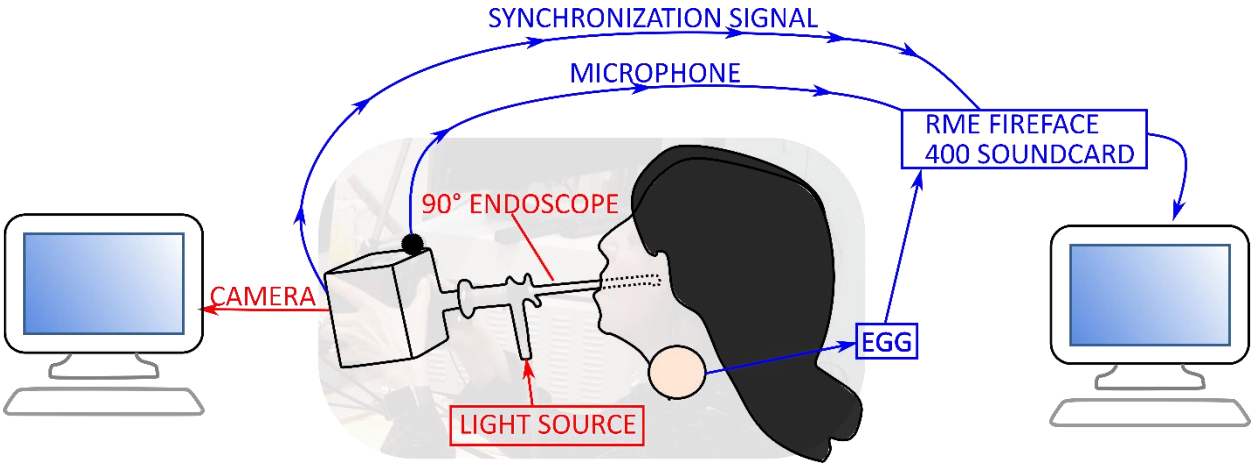


Figure 32. Experimental setup.

3.4.3.Results

On average, the outsiders were able to correctly identify the registers, as intended by the singer, in 64% of the cases, and the insiders in 89% of the cases. Results from the visual analysis of the high-speed videos are summarized in Figure 33. They revealed laryngeal and physiological differences separating the low-, middle-, and high-pitched phonation regions (indicated by R_{Lo}, R_{Mid}, and R_{Hi}, respectively). Importantly, the values of NAQ_{GA} and of the CgQ_{kym} were consistently lower in chest-like register throughout the entire range (except at B5 for the CgQ_{kym}), as shown in Figure 34. This indicates that the closing phase was consistently shorter in chest-like register.

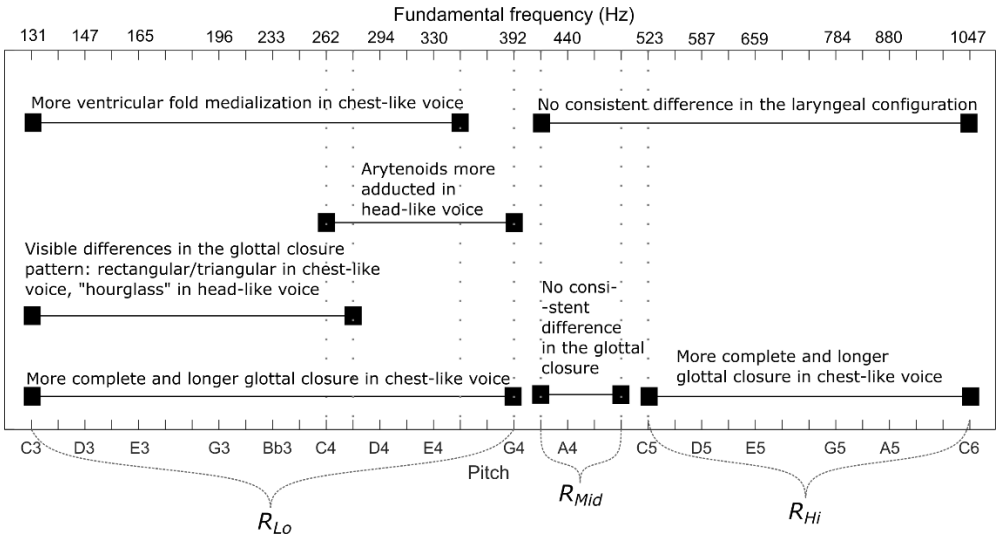


Figure 33. Diagram summarizing the visual and objective differences observed between the chest-like and head-like phonations in the high-speed videos and the corresponding ranges where those differences appeared. R_{Lo}, R_{Mid} and R_{Hi} indicate the low, middle, and high range, respectively.

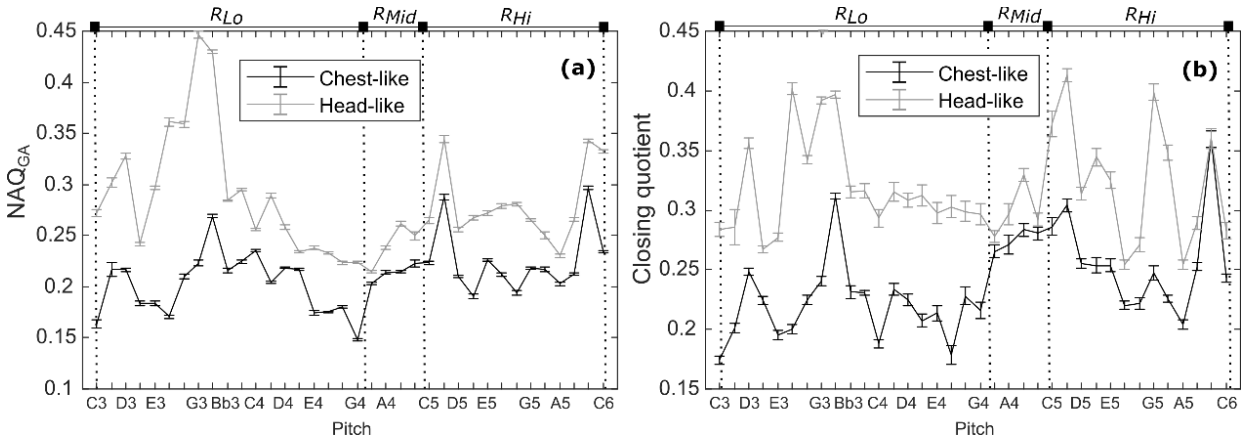


Figure 34. Mean Normalized Amplitude Quotient NAQ_{GA} (a) and Closing Quotient CgQ_{kym} (b) distinguishing the chest-like and head-like registers for every pitch. The error bars represent the standard error of the mean.

3.4.4.Discussion and conclusion

NAQ_{GA} and the CgQ_{kym} were found to be the only two parameters with consistently lower values in chest-like register across the entire pitch range (except at B5 for the CgQ_{kym}), which suggests the glottal closing speed is one of the most relevant parameters to discriminate between chest-like and head-like registers in our singer. This supports the hypothesis mentioned in section 2.4.2 that the increased activity of the TA muscle thickens the VFs and therefore increases the vibrational amplitude of the lower margin of the VFs, which in turn increases the closing speed. Interestingly, the CQ_{PVG} was not able to distinguish the two registers across the large pitch range (not shown here for brevity – see paper **III** for details). This conflicts with the results from previous studies that mostly used CQ and OQ for differentiating registers (Sundberg et al., 2001; Roubeau et al., 2009; Echternach, Burk, Köberlein, Herbst, et al., 2017; Echternach, Burk, Köberlein, Selamtzis, et al., 2017).

Perceptually, the lower score of the outsiders group in the listening tests indicates that those listeners had more difficulties than the insiders group to recognize the registers. This suggests that the distinction between the two registers intended by the singer is not easy to recognize. As expected, the insiders group performed better, as they were familiar with the study design.

*3.5.Paper **IV**: Simulation of laryngoscopic high-speed videos from a kinematic model*

3.5.1.Objectives

Clinicians commonly use laryngoscopic imaging techniques to assess the general VF health and the presence of pathologies. Particularly, strobolaryngoscopy, and to some extent, in some places, also videokymography (VKG), are routinely used, as they provide real-time feedback and their clinical value has been demonstrated (Casiano et al., 1992; Phadke et al., 2017). Laryngeal HSV, however, requires the recorded images to be reviewed and analyzed and is therefore slow to be routinely used in clinical practice (see also the reasons described in section 2.2.1.3), even though it is the technique that can provide the most detailed information about the VFs. As such, the relation between visual features observed in laryngeal HSV images and kinematic properties of the VFs is under researched, and there is a need to improve our interpretation of laryngeal HSV. This study proposes to address this need by extending a kinematic mucosal wave model of the VF vibrations to enable the generation of synthetic HSV of the entire glottis.

3.5.2.Methods

The mucosal wave model used in this study is described in section 2.6. This model was originally designed to simulate two-dimensional (2D) vibrations, by considering a coronal (vertical left-right) slice of the VFs. Here, it is extended to three-dimensional (3D) vibrations by concatenating 256 slices along the anterior-posterior axis. Kymograms generated by each slice are then combined, to form a simulated high-speed video, where the entire glottis is visible. Geometric and kinematic parameters of the VFs are then varied along the anterior-posterior axis, to display anterior-posterior differences typically observed in vivo through laryngeal HSV. Parameters were adjusted empirically so that the simulated videos resemble reference laryngeal high-speed videos, obtained from the Vienna database of pathological and non-pathological voices (Aichinger et al., 2016).

3.5.3.Results

For brevity, this section only describes two examples of HSV simulated by the model – see paper **IV** for details. The first example is displayed in Figure 35 and corresponds to a regular pressed phonation. For this example, the pressed quality is simulated by setting a negative glottal half-width across the entire glottis, except for the most posterior position where the glottal half-width is set to 0. The amplitude of vibration is also set to 0 at the posterior position, so that the left and right VFs are always only just touching each other at this position (this creates the impression of a slight posterior glottal gap in Figure 35c due to the roundedness of the VF shape). During glottal closing, the mucosal wave is seen propagating across the top VF surface (Figure 35a,b).

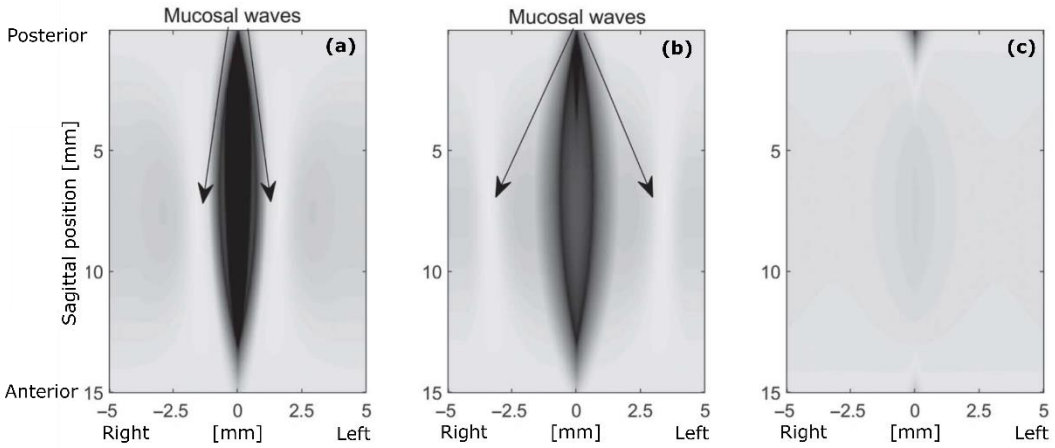


Figure 35. Illustration of pressed VF vibrations. Shown are frames of maximal opening (a), maximal closure (c), and a frame in between (b). The mucosal waves traveling laterally are also indicated by black arrows.

The second example corresponds to a pathological voice exhibiting chaotic vibrations, simulated by imposing random variations in the phase of vibration, which are different between the left and right VFs, and between their anterior and posterior part. This example is displayed through the PVG (recall section 2.2.1.4) in Figure 36. The vibration of both the left and right VFs are visibly irregular, in addition to being asymmetric.

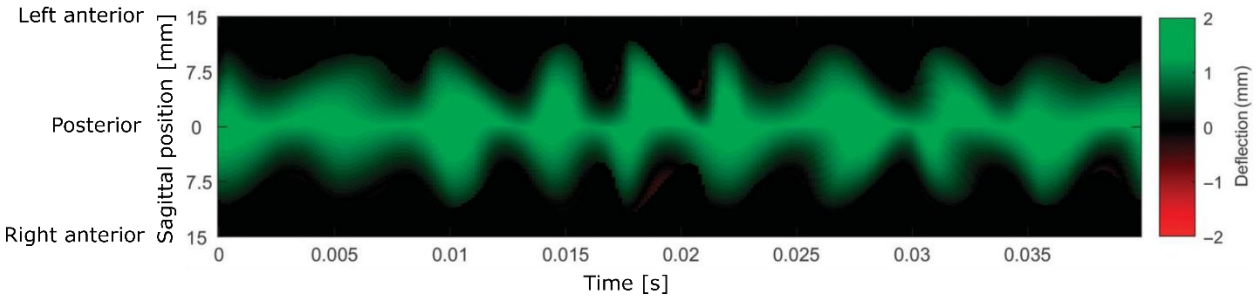


Figure 36. Phonovibrogram (PVG) corresponding to chaotic vibrations

3.5.4.Discussion and conclusion

The mucosal wave model is successfully extended to 3D, allowing the generation of synthetic videos with a realistic appearance, containing left-right and anterior-posterior differences. Additionally, the implementation of perturbations, in the form of random vibratory phase distortion, allows to synthetize irregular vibrations, which are associated with pathological voices. This includes perturbations that are different for the left and right VF, as well as for the anterior and posterior parts of the glottis. As such, this extended model offers useful insights about the effects of such perturbations in the VF kinematic parameters on the appearance of the corresponding vibrations.

4. OVERALL CONCLUSION

This thesis provides important new insights on the vibratory properties of the VFs. Specifically, the first part tests and documents a newly developed anechoic subglottal tract, to investigate the subglottal pressures and the VF vibrations in fully anechoic conditions, and to compare them with subglottally resonant conditions. It shows that the anechoic tract successfully suppresses its acoustic resonances. This subglottal tract is then used with excised deer larynges without a supraglottal tract to investigate the VF vibrations in fully anechoic conditions. In such conditions, the subglottal pressure waveform demonstrated a remarkable resemblance with the theoretical source flow waveform, which shows that the anechoic subglottal tract can be used to study the source signal without any acoustic interactions with resonance cavities and without the need for inverse filtering.

In the second part, this thesis investigates the differences in the occurrences of frequency jumps in human excised larynges, between anechoic and subglottally resonant conditions. This part presents the first experimental evidence that frequency jumps occur also in fully anechoic conditions. The results provide a proof that those frequency jumps are primarily caused by inherent nonlinear-dynamic VF properties, even though the acoustic resonances in the subglottal tract appear to play a secondary role.

The third part of this thesis investigates the relevant glottal and laryngeal parameters that can discriminate between chest and head registers throughout a wide pitch range. Here, chest and head registers are investigated over a wide pitch range in a professional female singer. Visual and objective analysis revealed that the glottal closing speed is one of the most relevant parameters to discern chest and head registers regardless of pitch, which is revealed through the consistent lower values of the normalized amplitude quotient and closing quotient in chest register. This contrasts with most previous studies that rather assumed the closed quotient to be the most sensitive measure for distinguishing the chest and head registers in singing. In our study, the closed quotient was not found to be able to distinguish the two registers as consistently as the normalized amplitude quotient and closing quotient. This part helps to advance our understanding of voice registers, particularly of the laryngeal adjustments and VF kinematics employed by trained singers to produce chest and head registers.

Finally, the fourth part aims at generating synthetic high-speed videos of the entire glottis. This was achieved through the extension of a two-dimensional existing kinematic model of the VF vibrations, referred to as the mucosal wave model here, to three dimensions. The extended model demonstrates capabilities of generating videos of the vibrating glottis that closely resemble the high-speed laryngoscopic videos obtained clinically in vivo. Particularly, the synthetic videos display left-right and anterior-posterior differences, which are typically seen in laryngeal voice disorders. Additionally, the extended model is capable of applying perturbations in the form of random phase distortions, which appear as irregular or chaotic vibrations. As such, this model could be further used to help understand the underlying mechanisms responsible for pathological voices.

5. REFERENCES

Aichinger, P., Roesner, I., Leonhard, M., Denk-Linnert, D.-M., Bigenzahn, W., & Schneider-Stickler, B. (2016). *A database of laryngeal high-speed videos with simultaneous high-quality audio recordings of pathological and non-pathological voices.* In Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16), Portorož, Slovenia.

Alberti, P. W. (1996). The History of Laryngology: A Centennial Celebration. *Otolaryngology–Head and Neck Surgery, 114*(3), 345-354.

Alderson, R. M. (1979). *Complete handbook of voice training*. West Nyack, New York: Parker Publishing Company, Inc.

Alipour, F., Finnegan, E. M., & Jaiswal, S. (2013). Phonatory Characteristics of the Excised Human Larynx in Comparison to Other Species. *J. Voice, 27*(4), 441-447.

Alku, P. (2011). Glottal inverse filtering analysis of human voice production—A review of estimation and parameterization methods of the glottal excitation and their applications. *Sadhana - Academy Proceedings in Engineering Sciences, 36*(5), 623-650.

Alku, P., Bäckström, T., & Vilkman, E. (2002). Normalized amplitude quotient for parametrization of the glottal flow. *J. Acoust. Soc. Am., 112*(2), 701-710.

Baer, T. (1975). *Investigation of phonation using excised larynxes (Doctoral dissertation).* Massachusetts Institute of Technology, Cambridge, MA.

Baer, T., Gay, T., & Niimi, S. (1976). Control of fundamental frequency, intensity and register of phonation. *Haskins Laboratories Status Report on Speech Research, 45*(46), 175-185.

Baer, T., Gore, J. C., Gracco, L. C., & Nye, P. W. (1991). Analysis of vocal tract shape and dimensions using magnetic resonance imaging: Vowels. *J. Acoust. Soc. Am., 90*(2), 799-828.

Baken, R. J. (1992). Electroglottography. *J. Voice, 6*(2), 98-110.

Baken, R. J., & Orlikoff, R. F. (2000). *Clinical measurement of speech and voice*: Cengage Learning.

Berke, G., Mendelsohn, A. H., Scott Howard, N., & Zhang, Z. (2013). Neuromuscular induced phonation in a human ex vivo perfused larynx preparation. *J. Acoust. Soc. Am., 133*(2), EL114-EL117.

Berry, D. A. (2001). Mechanisms of modal and nonmodal phonation. *J. Phon., 29*(4), 431-450.

Berry, D. A., Herzel, H., Titze, I. R., & Krischer, K. (1994). Interpretation of biomechanical simulations of normal and chaotic vocal fold oscillations with empirical eigenfunctions. *J. Acoust. Soc. Am., 95*(6), 3595-3604.

Berry, D. A., Herzel, H., Titze, I. R., & Story, B. H. (1996). Bifurcations in excised larynx experiments. *J. Voice, 10*(2), 129-138.

Berry, D. A., Montequin, D. W., & Tayama, N. (2001). High-speed digital imaging of the medial surface of the vocal folds. *J. Acoust. Soc. Am., 110*(5), 2539-2547.

Berry, D. A., & Titze, I. R. (1996). Normal modes in a continuum model of vocal fold tissues. *J. Acoust. Soc. Am., 100*(5), 3345-3354.

Björklund, S., & Sundberg, J. (2016). Relationship between subglottal pressure and sound pressure level in untrained voices. *J. Voice, 30*(1), 15-20.

Botros, A. (2001). Frequency to Musical Note Converter. Retrieved from https://newt.phys.unsw.edu.au/music/note/

Bouhuys, A., Mead, J., Proctor, D. F., & Stevens, K. N. (1968). Pressure-flow events during singing. *Ann. N. Y. Acad. Sci., 155*(1), 165-176.

Camacho, A., & Harris, J. G. (2008). A sawtooth waveform inspired pitch estimator for speech and music. *J. Acoust. Soc. Am., 124*(3), 1638-1652.

Casiano, R. R., Zaveri, V., & Lundy, D. S. (1992). Efficacy of Videostroboscopy in the Diagnosis of Voice Disorders. *Otolaryngology–Head and Neck Surgery, 107*(1), 95-100.

Castellengo, M., Lamesch, S., & Henrich, N. (2007). *Vocal registers and laryngeal mechanisms, a case study: The french «voix mixte».* In Proceedings of the 19th ICA, Madrid.

Chan, R. W., & Titze, I. R. (2003). Effect of Postmortem Changes and Freezing on the Viscoelastic Properties of Vocal Fold Tissues. *Annals of Biomedical Engineering, 31*(4), 482-491.

Collyer, S., Davis, P. J., Thorpe, C. W., & Callaghan, J. (2007). Sound pressure level and spectral balance linearity and symmetry in the messa di voce of female classical singers. *J. Acoust. Soc. Am., 121*(3), 1728-1736.

Collyer, S., Davis, P. J., Thorpe, C. W., & Callaghan, J. (2009). Fundamental frequency influences the relationship between sound pressure level and spectral balance in female classically trained singers. *J. Acoust. Soc. Am., 126*(1), 396-406.

Colton, R. H. (1972). Spectral characteristics of the modal and falsetto registers. *Folia Phoniatr. Logop., 24*(5-6), 337-344.

Cranen, B., & Boves, L. (1987). On subglottal formant analysis. *J. Acoust. Soc. Am., 81*(3), 734-746.

Deliyski, D. D., Hillman, R. E., & Mehta, D. D. (2015). Laryngeal High-Speed Videoendoscopy: Rationale and Recommendation for Accurate and Consistent Terminology. *J. Speech Lang. Hear. Res., 58*(5), 1488-1492.

Deliyski, D. D., Powell, M. E. G., Zacharias, S. R. C., Gerlach, T. T., & de Alarcon, A. (2015). Experimental Investigation on Minimum Frame Rate Requirements of High-Speed Videoendoscopy for Clinical Voice Assessment. *Biomed Signal Process Control, 17*, 21-28.

Döllinger, M., & Berry, D. A. (2006). Visualization and quantification of the medial surface dynamics of an excised human vocal fold during phonation. *J. Voice, 20*(3), 401-413.

Döllinger, M., Kobler, J., Berry, D. A., Mehta, D. D., Luegmair, G., & Bohr, C. (2011). Experiments on Analysing Voice Production: Excised (Human, Animal) and In Vivo (Animal) Approaches. *Curr. Bioinform., 6*(3), 286-304.

Döllinger, M., Lohscheller, J. r., McWhorter, A., & Kunduk, M. (2009). Variability of normal vocal fold dynamics for different vocal loading in one healthy subject investigated by phonovibrograms. *J. Voice, 23*(2), 175-181.

Durham, P. L., Scherer, R. C., Druker, D. G., & Titze, I. R. (1987). Development of excised larynx procedures for studying mechanisms of phonation. Technical report: Voice Acoustics and Biomechanics Laboratory, Department of Speech Pathology and Audiology, The University of Iowa, IA, USA.

Echternach, M., Burk, F., Köberlein, M., Herbst, C. T., Döllinger, M., Burdumy, M., & Richter, B. (2017). Oscillatory characteristics of the vocal folds across the tenor passaggio. *J. Voice, 31*(3), 381.e385-381.e314.

Echternach, M., Burk, F., Köberlein, M., Selamtzis, A., Döllinger, M., Burdumy, M., . . . Herbst, C. T. (2017). Laryngeal evidence for the first and second passaggio in professionally trained sopranos. *PLoS One, 12*(5), e0175865.

Eysholdt, U., Tigges, M., Wittenberg, T., & Pröschel, U. (1996). Direct Evaluation of High-Speed Recordings of Vocal Fold Vibrations. *Folia Phoniatr. Logop., 48*(4), 163-170.

Fabre, P. (1957). Un procédé électrique percutané d'inscrition de l'accolement glottique au cours de la phonation: glottographie de haute fréquence. Premiers résultats. *Bull. Acad. Natl. Med., 141*, 66.

Fant, G. (1960). *Acoustic Theory of Speech Production* Mouton, The Hague.

Fant, G., Ishizaka, K., Lindqvist, J., & Sundberg, J. (1972). Subglottal formants. *STL-QPSR, 1*(1972), 1-12.

Farnsworth, D. W. (1940). High-speed motion pictures of the human vocal cords. *Bell Laboratories Record, 18*, 203-208.

Ferrein, A. (1741). De la formation de la voix de l'homme. *Histoire de l'Académie Royale des Sciences, 1*, 409-432.
Fitch, W. T., & Giedd, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *J. Acoust. Soc. Am., 106*(3), 1511-1522.
Flanagan, J. L. (1972). *Speech analysis: Synthesis and perception, 2nd ed*. Oxford, England: Springer-Verlag.
Flanagan, J. L., & Landgraf, L. L. (1968). Self-Oscillating Source for Vocal-Tract Synthesizers. *IEEE Transactions on Audio and Electroacoustics, AU16*(1), 57-&.
Garcia, M. (1847). *Mémoire sur la voix humaine présenté à l'Académie des Sciences en 1840*: Duverger.
Garcia, M., & Herbst, C. T. (2018). Excised larynx experimentation: history, current developments, and prospects for bioacoustic research. *Anthropol. Sci., 126*(1), 9-17.
Garnier, M., Henrich, N., Crevier-Buchman, L., Vincent, C., Smith, J., & Wolfe, J. (2012). Glottal behavior in the high soprano range and the transition to the whistle register. *J. Acoust. Soc. Am., 131*(1), 951-962.
Grillo, H. C., Dignan, E. F., Miura, T., & Scannell, J. G. (1964). Extensive resection and reconstruction of mediastinal trachea without prosthesis or graft: an anatomical study in man. *The Journal of Thoracic and Cardiovascular Surgery, 48*(5), 741-749.
Hampala, V., Garcia, M., Švec, J. G., Scherer, R. C., & Herbst, C. T. (2016). Relationship between the electroglottographic signal and vocal fold contact area. *J. Voice, 30*(2), 161-171.
Hampala, V., Švec, J. G., Schovánek, D., & Mandát, D. (2013). Užitný vzor č. 25585: Model subglotického traktu. [Utility model no. 25585: Model of subglottal tract.] Soukup, P. 2013-27834(CZ 25505 U1), 1-7. 24-6-2013. Praha, Úřad průmyslového vlastnictví. [Prague, Czech Republic, Industrial property office].
Henrich Bernardoni, N. (2006). Mirroring the voice from Garcia to the present day: some insights into singing voice registers. *Logoped. Phoniatr. Vocol., 31*(1), 3-14.
Henrich Bernardoni, N., D'Alessandro, C., Doval, B., & Castellengo, M. (2004). On the use of the derivative of electroglottographic signals for characterization of nonpathological phonation. *J. Acoust. Soc. Am., 115*(3), 1321-1332.
Henrich Bernardoni, N., d'Alessandro, C., Doval, B., & Castellengo, M. (2005). Glottal open quotient in singing: Measurements and correlation with laryngeal mechanisms, vocal intensity, and fundamental frequency. *J. Acoust. Soc. Am., 117*(3), 1417-1430.
Henrich Bernardoni, N., Roubeau, B., & Castellengo, M. (2003). *On the use of electroglottography for characterisation of the laryngeal mechanisms.* In Proceedings of SMAC 2003, Stockholm Music Acoustics Conference, Stockholm, Sweden.
Herbst, C. T. (2019). Electroglottography – An Update. *J. Voice*.
Herbst, C. T., Elemans, C. P. H., Tokuda, I. T., Chatziioannou, V., & Švec, J. G. (2023). Dynamic System Coupling in Voice Production. *J. Voice*.
Herbst, C. T., Hampala, V., Garcia, M., Hofer, R., & Švec, J. G. (2017). Hemi-laryngeal Setup for Studying Vocal Fold Vibration in Three Dimensions. *J. Vis. Exp.*(129).
Herbst, C. T., Lohscheller, J., Švec, J. G., Henrich, N., Weissengruber, G., & Fitch, W. T. (2014). Glottal opening and closing events investigated by electroglottography and super-high-speed video recordings. *J. Exp. Biol., 217*(6), 955-963.
Herbst, C. T., Nishimura, T., Garcia, M., Migimatsu, K., & Tokuda, I. T. (2020). Effect of Ventricular Folds on Vocalization Fundamental Frequency in Domestic Pigs (Sus scrofa domesticus). *J. Voice*.
Herbst, C. T., Qiu, Q., Schutte, H. K., & Švec, J. G. (2011). Membranous and cartilaginous vocal fold adduction in singing. *J. Acoust. Soc. Am., 129*(4), 2253-2262.
Herbst, C. T., & Švec, J. G. (2014). Adjustment of glottal configurations in singing. *J. Sing., 70*, 301-308.

Hertegård, S. (2005). What have we learned about laryngeal physiology from high-speed digital videoendoscopy? *Curr Opin Otolaryngol Head Neck Surg, 13*(3), 152-156.
Herzel, H., Berry, D. A., Titze, I. R., & Saleh, M. (1994). Analysis of Vocal Disorders With Methods From Nonlinear Dynamics. *J. Speech Lang. Hear. Res., 37*(5), 1008-1019.
Hirano, M. (1974). Morphological structure of the vocal cord as a vibrator and its variations. *Folia Phoniatr. Logop., 26*(2), 89-94.
Hirano, M. (1975). Phonosurgery: basic and clinical investigations. *Otologia (Fukuoka), 21 (suppl. 1)*, 239-440.
Hirano, M. (1981). *Clinical examination of voice* (Vol. 5): Spring-Verlag.
Hirano, M., Kakita, Y., Kawasaki, H., Gould, W. J., & Lambiase, A. (1981). Data from high-speed motion picture studies. In K. N. Stevens & M. Hirano (Eds.), *Vocal Fold Physiology* (pp. 85-93): Univ. Tokyo Press.
Hirano, M., Vennard, W., & Ohala, J. (1970). Regulation of register, pitch and intensity of voice. An electromyographic investigation of intrinsic laryngeal muscles. *Folia Phoniatr. Logop., 22*(1).
Hiroto, I. (1966). The mechanism of phonation: its pathophysiological aspects. (in Japanese). *Nippon Jibiinkoka Gakkai Kaiho, 69*(12), 2097-2106.
Holbert, J. M., & Strollo, D. C. (1995). Imaging of the Normal Trachea. *Journal of Thoracic Imaging, 10*(3).
Horáček, J., & Švec, J. G. (2002). Aeroelastic model of vocal-fold-shaped vibrating element for studying the phonation threshold. *J. Fluids Struct., 16*(7), 931-955.
Horáček, J., Švec, J. G., & Šidlof, P. (2009). *Numerical simulation of videokymographic images of self-oscillating vocal folds.* In 3rd Advanced Voice Function Assessment (AFVA) International Workshop.
Horáček, J., Švec, J. G., Veselý, J., & Vilkman, E. (2004). *Bifurcations in Excised Larynges Caused by Vocal Fold Elongation.* In Proceedings of the International Conference on Voice Physiology and Biomechanics, Marseille: Laboratory of Audio-Phonology, France.
Husson, R. (1950). *Etude des phénomènes physiologiques et acoustiques fondamentaux de la voix chantée [Study of the fundamental physiologic and acoustic phenonema of the singing voice] (Doctoral thesis).* Faculté des Sciences, Éditions de la revue scientifique, Paris.
Ishizaka, K., & Flanagan, J. L. (1972). Synthesis of Voiced Sounds From a Two-Mass Model of the Vocal Cords. *Bell Syst. tech., 51*(6), 1233-1268.
Jiang, J. J., Chang, C. I., Raviv, J. R., Gupta, S., Banzali Jr, F. M., & Hanson, D. G. (2000). Quantitative study of mucosal wave via videokymography in canine larynges. *Laryngoscope, 110*(9), 1567-1573.
Jiang, J. J., & Titze, I. R. (1993). A methodological study of hemilaryngeal phonation. *Laryngoscope, 103*(8), 872-882.
Jiang, J. J., Zhang, Y., Kelly, M. P., Bieging, E. T., & Hoffman, M. R. (2008). An automatic method to quantify mucosal waves via videokymography. *Laryngoscope, 118*(8), 1504-1510.
Kallen, L. A. (1932). LARYNGOSTROBOSCOPY IN THE PRACTICE OF OTOLARYNGOLOGY. *Archives of Otolaryngology, 16*(6), 791-807.
Karakozoglou, S.-Z., Henrich, N., d'Alessandro, C., & Stylianou, Y. (2012). Automatic glottal segmentation using local-based active contours and application to glottovibrography. *Speech Commun., 54*(5), 641-654.
Kinsler, L. E., Frey, A. R., Coppens, A. B., & Sanders, J. V. (2000). *Fundamentals of acoustics (4th edition)*. New York: John Wiley & Sons, Inc.
Kist, A. M., Gómez, P., Dubrovskiy, D., Schlegel, P., Kunduk, M., Echternach, M., . . . Döllinger, M. (2021). A Deep Learning Enhanced Novel Software Tool for Laryngeal Dynamics Analysis. *J. Speech Lang. Hear. Res., 64*(6), 1889-1903.
Kist, A. M., Zilker, J., Gómez, P., Schützenberger, A., & Döllinger, M. (2020). Rethinking glottal midline detection. *Sci. Rep., 10*.

Kob, M., & Frauenrath, T. (2009). A system for parallel measurement of glottis opening and larynx position. *Biomed Signal Process Control, 4*(3), 221-228.

Kochis-Jennings, K. A., Finnegan, E. M., Hoffman, H. T., & Jaiswal, S. (2012). Laryngeal muscle activity and vocal fold adduction during chest, chestmix, headmix, and head registers in females. *J. Voice, 26*(2), 182-193.

Kochis-Jennings, K. A., Finnegan, E. M., Hoffman, H. T., Jaiswal, S., & Hull, D. (2014). Cricothyroid muscle and thyroarytenoid muscle dominance in vocal register control: preliminary results. *J. Voice, 28*(5), 652.e621-652.e629.

Krausert, C. R., Olszewski, A. E., Taylor, L. N., McMurray, J. S., Dailey, S. H., & Jiang, J. J. (2011). Mucosal Wave Measurement and Visualization Techniques. *J. Voice, 25*(4), 395-405.

Kreiman, J., Gerratt, B. R., & Antoñanzas-Barroso, N. (2007). Measures of the Glottal Source Spectrum. *J. Speech Lang. Hear. Res., 50*(3), 595-610.

Kumar, S. P., Phadke, K. V., Vydrová, J., Novozámský, A., Zita, A., Zitová, B., & Švec, J. G. (2020). Visual and Automatic Evaluation of Vocal Fold Mucosal Waves Through Sharpness of Lateral Peaks in High-Speed Videokymographic Images. *J. Voice, 34*(2), 170-178.

Kumar, S. P., & Švec, J. G. (2019). Kinematic model for simulating mucosal wave phenomena on vocal folds. *Biomed Signal Process Control, 49*, 328-337.

Lagier, A., Guenoun, D., Legou, T., Espesser, R., Giovanni, A., & Champsaur, P. (2017). Control of the glottal configuration in ex vivo human models: quantitative anatomy for clinical and experimental practices. *Surg. Radiol. Anat., 39*(3), 257-262.

Lagier, A., Legou, T., Silva, F., & Hélie, T. (2020). *Investigating phonation through excised human larynges: recent developments and ongoing works on an animated testbench*. Paper presented at the eForum Acusticum 2020.

Large, J. (1972). Towards an integrated physiologic-acoustic theory of vocal registers. *NATS Bulletin, 28*(3), 18-25.

Legou, T., Lagier, A., Silva, F., Bernardoni, N. H., Champsaur, P., & Giovanni, A. (2015). *Test bench for human excised larynx studies.* In 9th International Workshop Models and Analysis of Vocal Emissions for Biomedical Applications (MAVEBA 2015).

Lohscheller, J., & Eysholdt, U. (2008). Phonovibrogram visualization of entire vocal fold dynamics. *Laryngoscope, 118*(4), 753-758.

Lulich, S. M., Morton, J. R., Arsikere, H., Sommers, M. S., Leung, G. K. F., & Alwan, A. (2012). Subglottal resonances of adult male and female native speakers of American English. *J. Acoust. Soc. Am., 132*(4), 2592-2602.

Mehta, D. D., Deliyski, D. D., & Hillman, R. E. (2010). Commentary on Why Laryngeal Stroboscopy Really Works: Clarifying Misconceptions Surrounding Talbot's Law and the Persistence of Vision. *J. Speech Lang. Hear. Res., 53*(5), 1263-1267.

Merkel, C. L. (1863). Anatomie und Physiologie des menschlichen Stimm-und Sprachprogramms. *Antropophonik. Leipzig: Abel*.

Miller, D. G., & Schutte, H. K. (1985). *Characteristic patterns of sub-and supraglottal pressure variations within the glottal cycle.* In Transcr. XIIIth Symp. Care Prof. Voice.

Moore, P. (1991). A short history of laryngeal investigation. *J. Voice, 5*(3), 266-281.

Mörner, M., Fransson, F., & Fant, G. (1963). Voice register terminology and standard pitch. *STL-QPSR, 4*(4), 17-23.

Murtola, T., Aalto, A., Malinen, J., Aalto, D., & Vainio, M. (2018). Modal locking between vocal fold oscillations and vocal tract acoustics. *Acta Acust. united Ac., 104*(2), 323-337.

Novozámský, A., Sedlář, J., Zita, A., Šroubek, F., Flussef, J., Švec, J. G., . . . Zitová, B. (2015). *Image analysis of videokymographic data.* In 2015 IEEE International Conference on Image Processing (ICIP).

Patel, R. R., & Ternström, S. (2021). Quantitative and Qualitative Electroglottographic Wave Shape Differences in Children and Adults Using Voice Map–Based Analysis. *J. Speech Lang. Hear. Res., 64*(8), 2977-2995.

Phadke, K. V., Vydrová, J., Domagalská, R., & Švec, J. G. (2017). Evaluation of clinical value of videokymography for diagnosis and treatment of voice disorders. *Eur. Arch. Oto-Rhino-Laryngol., 274*(11), 3941-3949.

Qiu, Q., & Schutte, H. K. (2007). Real-time kymographic imaging for visualizing human vocal-fold vibratory function. *Review of Scientific Instruments, 78*(2), 024302.

Rothenberg, M. (1973). A new inverse-filtering technique for deriving the glottal air flow waveform during voicing. *J. Acoust. Soc. Am., 53*(6), 1632-1645.

Rothenberg, M. (1981). Acoustic interaction between the glottal source and the vocal tract. In *Vocal fold physiology* (Vol. 1, pp. 305-323).

Rothenberg, M. (1992). A multichannel electroglottograph. *J. Voice, 6*(1), 36-43.

Roubeau, B., Chevrie-Muller, C., & Arabia-Guidet, C. (1987). Electroglottographic study of the changes of voice registers. *Folia Phoniatr. Logop., 39*(6), 280-289.

Roubeau, B., Henrich, N., & Castellengo, M. (2009). Laryngeal Vibratory Mechanisms: The Notion of Vocal Register Revisited. *J. Voice, 23*(4), 425-438.

Sadolin, C. (2000). *Complete vocal technique*: Shout Publishing Copenhagen, Denmark.

Scherer, R. C., Druker, D. G., & Titze, I. R. (1988). Electroglottography and Direct Measurement of Vocal Fold Contact Area. In O. Fujimura (Ed.), *Vocal fold physiology: voice production, mechanisms, and functions* (pp. 279-291). New York: Raven Press Ltd.

Scherer, R. C., Shinwari, D., De Witt, K. J., Zhang, C., Kucinschi, B. R., & Afjeh, A. A. (2001). Intraglottal pressure profiles for a symmetric and oblique glottis with a divergence angle of 10 degrees. *J. Acoust. Soc. Am., 109*(4), 1616-1630.

Schlegel, P., Stingl, M., Kunduk, M., Kniesburges, S., Bohr, C., & Döllinger, M. (2019). Dependencies and Ill-designed Parameters Within High-speed Videoendoscopy and Acoustic Signal Analysis. *J. Voice, 33*(5), 811.e811-811.e812.

Schoentgen, J., & Aichinger, P. (2015). *Synthetic kymograms and glottal area waveforms in simulated non-neutral phonation.* In 9th International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications, MAVEBA 2015.

Schutte, H. K., & Miller, D. G. (1988). Resonanzspiele der Gesangsstimme in ihren Beziehungen zu supra- und subglottalen Druckverlaufen: Konsequenzen für die Stimmbildungstheorie. [Play of Resonances in the Singing Voice in the Supra- and Subglottal Pressure Changes: Consequences for the Theory of Voice Production]. *Folia Phoniatr. Logop., 40*, 65-73.

Stark, J. (1999). *Bel canto: a history of vocal pedagogy*. Toronto/Buffalo/London: University of Toronto Press.

Story, B. H., Titze, I. R., & Hoffman, E. A. (1996). Vocal tract area functions from magnetic resonance imaging. *J. Acoust. Soc. Am., 100*(1), 537-554.

Sundberg, J. (2018). Flow glottogram and subglottal pressure relationship in singers and untrained voices. *J. Voice, 32*(1), 23-31.

Sundberg, J. (2022). Objective Characterization of Phonation Type Using Amplitude of Flow Glottogram Pulse and of Voice Source Fundamental. *J. Voice, 36*(1), 4-14.

Sundberg, J., Andersson, M., & Hultqvist, C. (1999). Effects of subglottal pressure variation on professional baritone singers' voice sources. *J. Acoust. Soc. Am., 105*(3), 1965-1971.

Sundberg, J., & Högset, C. (2001). Voice source differences between falsetto and modal registers in counter tenors, tenors and baritones. *Logoped. Phoniatr. Vocol., 26*(1), 26-36.

Sundberg, J., Scherer, R., Hess, M., Müller, F., & Granqvist, S. (2013). Subglottal pressure oscillations accompanying phonation. *J. Voice, 27*(4), 411-421.

Švancara, P., Horáček, J., Martínek, T., & Švec, J. G. (2014). *Numerical simulation of videokymographic images from the results of the finite element model.* In Engineering Mechanics.
Švec, J. G., & Granqvist, S. (2018). Tutorial and Guidelines on Measurement of Sound Pressure Level in Voice and Speech. *J. Speech Lang. Hear. Res., 61*(3), 441-461.
Švec, J. G., & Schutte, H. K. (1996). Videokymography: high-speed line scanning of vocal fold vibration. *J. Voice, 10*(2), 201-205.
Švec, J. G., Schutte, H. K., Chen, J. C., & Titze, I. R. (2021). Integrative Insights into the Myoelastic-Aerodynamic Theory and Acoustics of Phonation. Scientific Tribute to Donald G. Miller. *J. Voice*.
Švec, J. G., Schutte, H. K., & Miller, D. G. (1999). On pitch jumps between chest and falsetto registers in voice: Data from living and excised human larynges. *J. Acoust. Soc. Am., 106*(3), 1523-1531.
Švec, J. G., & Šram, F. (2002). *Kymographic imaging of the vocal fold oscillations*. Paper presented at the 7th International Conference on Spoken Language Processing, Denver, Colorado.
Švec, J. G., Šram, F., & Schutte, H. K. (2007). Videokymography in Voice Disorders: What to Look For? *Ann Otol Rhinol Laryngol, 116*(3), 172-180.
Švec, J. G., Šram, F., & Schutte, H. K. (2009). Videokymography. In *The larynx, Third Edition, Volume 1* (pp. 251-271): Plural Publishing, San Diego.
Švec, J. G., Sundberg, J., & Hertegard, S. (2008). Three registers in an untrained female singer analyzed by videokymography, strobolaryngoscopy and sound spectrography. *J. Acoust. Soc. Am., 123*(1), 347-353.
Ternström, S. (2019). Normalized time-domain parameters for electroglottographic waveforms. *J. Acoust. Soc. Am., 146*(1), EL65-EL70.
Titze, I. R. (1980). Comments on the myoelastic-aerodynamic theory of phonation. *J. Speech Lang. Hear. Res., 23*(3), 495-510.
Titze, I. R. (1988a). A framework for the study of vocal registers. *J. Voice, 2*(3), 183-194.
Titze, I. R. (1988b). The physics of small-amplitude oscillation of the vocal folds. *J. Acoust. Soc. Am., 83*(4), 1536-1552.
Titze, I. R. (1989). A four-parameter model of the glottis and vocal fold contact area. *Speech Commun., 8*(3), 191-201.
Titze, I. R. (1990). Interpretation of the electroglottographic signal. *J. Voice, 4*(1), 1-9.
Titze, I. R. (2000a). Basic anatomy of the larynx. In *Principles of Voice Production (second printing)* (pp. 1-22): National Center for Voice and Speech Iowa City IA 52242.
Titze, I. R. (2000b). *Principles of Voice Production (second printing)*: National Center for Voice and Speech Iowa City IA 52242.
Titze, I. R. (2000c). The Source-filter theory of vowels. In *Principles of Voice Production (second printing)* (pp. 149-184): National Center for Voice and Speech Iowa City IA 52242.
Titze, I. R. (2000d). Vocal fold oscillation. In *Principles of Voice Production (second printing)* (pp. 87-122): National Center for Voice and Speech Iowa City IA 52242.
Titze, I. R. (2006a). Experiments with excised larynges. In *The myoelastic aerodynamic theory of phonation* (pp. 1-62). Denver CO and Iowa City IA: National Center for Voice and Speech.
Titze, I. R. (2006b). *The Myoelastic Aerodynamic Theory of Phonation*. Iowa City lA 52242: National Center for Voice and Speech.
Titze, I. R. (2006c). Theoretical analysis of maximum flow declination rate versus maximum area declination rate in phonation. *J. Speech Lang. Hear. Res., 49*(2), 439-447.
Titze, I. R. (2008). Nonlinear source-filter coupling in phonation: theory. *J. Acoust. Soc. Am., 123*(5), 2733-2749.
Titze, I. R., Baken, R. J., Bozeman, K. W., Granqvist, S., Henrich, N., Herbst, C. T., . . . Wolfe, J. (2015). Toward a consensus on symbolic notation of harmonics, resonances, and formants in vocalization. *J. Acoust. Soc. Am., 137*(5), 3005-3007.

Titze, I. R., Jiang, J., & Drucker, D. G. (1988). Preliminaries to the body-cover theory of pitch control. *J. Voice, 1*(4), 314-319.
Titze, I. R., Jiang, J. J., & Hsiao, T.-Y. (1993). Measurement of mucosal wave propagation and vertical phase difference in vocal fold vibration. *Ann Otol Rhinol Laryngol, 102*(1), 58-63.
Titze, I. R., Luschei, E. S., & Hirano, M. (1989). Role of the thyroarytenoid muscle in regulation of fundamental frequency. *J. Voice, 3*(3), 213-224.
Titze, I. R., Riede, T., & Popolo, P. (2008). Nonlinear source–filter coupling in phonation: Vocal exercises. *J. Acoust. Soc. Am., 123*(4), 1902-1915.
Tokuda, I. T., Horáček, J., Švec, J. G., & Herzel, H. (2007). Comparison of biomechanical modeling of register transitions and voice instabilities with excised larynx experiments. *J. Acoust. Soc. Am., 122*(1), 519-531.
Tokuda, I. T., Horáček, J., Švec, J. G., & Herzel, H. (2008). Bifurcations and chaos in register transitions of excised larynx experiments. *Chaos Interdiscip. J. Nonlinear Sci., 18*(1), 013102.
Tokuda, I. T., Zemke, M., Kob, M., & Herzel, H. (2010). Biomechanical modeling of register transitions and the role of vocal tract resonators. *J. Acoust. Soc. Am., 127*(3), 1528-1536.
Van den Berg, J. (1958). Myoelastic-aerodynamic theory of voice production. *J. Speech Lang. Hear. Res., 1*, 227-244.
Van den Berg, J., & Tan, T. S. (1959). Results of Experiments with Human Larynxes. *Pract. Oto-Rhino-Laryng., 21*(6), 425-450.
Van den Berg, J., Vennard, W., Burger, D., & Shervanian, C. C. (Writers). (1960). Voice production. The vibrating larynx. (Instructional film). In. Groningen, the Netherlands: University of Groningen.
Veltrup, R., Angerer, S., Semmler, M., Kist, A. M., Zillig, T., Zilker, J., . . . Döllinger, M. (2021). *Multi-Modal Ex Vivo Setup For 3d Imaging Of The Medial And Superior Vocal Fold Surfaces In Hemi Larynges.* In 14th International Conference Advances In Quantitative Laryngology, Voice And Speech Research (AQL).
Verdolini, K., & Titze, I. R. (1994). The Application of Laboratory Formulas to Clinical Voice Management. *National Center for Voice and Speech Status and Progress Report, 7*, 197-205.
Verikas, A., Uloza, V., Bacauskiene, M., Gelzinis, A., & Kelertas, E. (2009). Advances in laryngeal imaging. *Eur. Arch. Oto-Rhino-Laryngol., 266*(10), 1509-1520.
Vilkman, E. (1987). An Apparatus for Studying the Role of the Cricothyroid Articulation in the Voice Production of Excised Human Larynges. *Folia Phoniatr. Logop., 39*(4), 169-177.
Vorperian, H. K., Wang, S., Chung, M. K., Schimek, E. M., Durtschi, R. B., Kent, R. D., . . . Gentry, L. R. (2009). Anatomic development of the oral and pharyngeal portions of the vocal tract: An imaging study. *J. Acoust. Soc. Am., 125*(3), 1666-1678.
Wade, L., Hanna, N., Smith, J., & Wolfe, J. (2016). *Soprano singing, with and without resonances.* In Proceedings of Meetings on Acoustics.
Wade, L., Hanna, N., Smith, J., & Wolfe, J. (2017). The role of vocal tract and subglottal resonances in producing vocal instabilities. *J. Acoust. Soc. Am., 141*(3), 1546.
Webb, M. (2008). Mal Webb Sideways Yodeling. Retrieved from https://www.youtube.com/watch?v=8_6nNWX7TTI
Wolfe, J. (2001). Note names, MIDI numbers and frequencies. Retrieved from https://newt.phys.unsw.edu.au/jw/notes.html
Wolfe, J., Tze Wei Chu, D., Chen, J.-M., & Smith, J. (2016). An Experimentally Measured Source–Filter Model: Glottal Flow, Vocal Tract Gain and Output Sound from a Physical Model. *Acoust. Aust., 44*(1), 187-191.
Wu, L., & Zhang, Z. (2019). Voice production in a MRI-based subject-specific vocal fold model with parametrically controlled medial surface shape. *J. Acoust. Soc. Am., 146*(6), 4190-4198.

Yamauchi, A., Yokonishi, H., Imagawa, H., Sakakibara, K.-I., Nito, T., Tayama, N., & Yamasoba, T. (2015). Quantitative Analysis of Digital Videokymography: A Preliminary Study on Age- and Gender-Related Difference of Vocal Fold Vibration in Normal Speakers. *J. Voice, 29*(1), 109-119.

Zañartu, M., Mehta, D. D., Ho, J. C., Wodicka, G. R., & Hillman, R. E. (2011). Observation and analysis of in vivo vocal fold tissue instabilities produced by nonlinear source-filter coupling: a case study. *J. Acoust. Soc. Am., 129*(1), 326-339.

Zemlin, W. R. (1981). Phonation. In W. R. Zemlin (Ed.), *Speech and hearing science: anatomy and physiology (second edition)* (pp. 100-196). Englewood Cliffs: Prentice-Hall, Inc.

Zhang, Z. (2016). Cause-effect relationship between vocal fold physiology and voice production in a three-dimensional phonation model. *J. Acoust. Soc. Am., 139*(4), 1493-1507.

Zhang, Z., Neubauer, J., & Berry, D. A. (2006). The influence of subglottal acoustics on laboratory models of phonation. *J. Acoust. Soc. Am., 120*(3), 1558-1569.

Zita, A., Novozámský, A., Zitová, B., Šorel, M., Herbst, C. T., Vydrová, J., & Švec, J. G. (2022). Videokymogram Analyzer Tool: Human–computer comparison. *Biomed Signal Process Control, 78*, 103878.

Supplement A: Paper I

Subglottal pressure oscillations in anechoic and resonant conditions and their influence on excised larynx phonations.

H. LEHOUX, V. HAMPALA, J. G. ŠVEC

Scientific Reports, **11**(1), 28 (2021)

doi: 10.1038/s41598-020-79265-3

scientific reports

Check for updates

OPEN

Subglottal pressure oscillations in anechoic and resonant conditions and their influence on excised larynx phonations

Hugo Lehoux, Vít Hampala & Jan G. Švec✉

Excised larynges serve as natural models for studying behavior of the voice source. Acoustic resonances inside the air-supplying tubes below the larynx (i.e., subglottal space), however, interact with the vibratory behavior of the larynges and obscure their inherent vibration properties. Here, we explore a newly designed anechoic subglottal space which allows removing its acoustic resonances. We performed excised larynx experiments using both anechoic and resonant subglottal spaces in order to analyze and compare, for the very first time, the corresponding subglottal pressures, electroglottographic and radiated acoustic waveforms. In contrast to the resonant conditions, the anechoic subglottal pressure waveforms showed negligible oscillations during the vocal fold contact phase, as expected. When inverted, these waveforms closely matched the inverse filtered radiated sound waveforms. Subglottal resonances modified also the radiated sound pressures (Level 1 interactions). Furthermore, they changed the fundamental frequency (f_o) of the vocal fold oscillations and offset phonation threshold pressures (Level 2 interactions), even for subglottal resonance frequencies 4–10 times higher than f_o. The obtained data offer the basis for better understanding the inherent vibratory properties of the vocal folds, for studying the impact of structure-acoustic interactions on voice, and for validation of computational models of voice production.

The well-known source-filter theory proposed by Gunnar Fant[1] described the voice production mechanism as a sound source (the exhalatory air flow modulated by vocal fold vibrations), filtered by acoustic resonances in the supraglottal cavities above the vocal folds (i.e. vocal tract resonances), and supposed no interaction between the source and the filter. Although this theory works well for speech analysis and synthesis, it is not sufficient to explain several voice production phenomena, such as vocal fold self-oscillations and voice instabilities. Deeper insights into the mechanism of the self-sustained vocal fold oscillation are provided through the myoelastic-aerodynamic (MEAD) theory of voice production formulated by van den Berg[2] and further elaborated by Titze[3–5]. This theory predicts interactions between the vocal fold vibrations and the surrounding pressures (i.e., subglottal and supraglottal pressures), leading to interdependency and nonlinear phenomena[6–8]. Titze categorized the source-filter interactions in two levels: Level 1 interactions exhibit changes in the source flow waveform; Level 2 interactions exhibit changes in the vocal fold oscillations[9]. Observations on the interaction phenomena preceded the MEAD theory: as early as in 1932, D. Weiss reported on singing voice instabilities induced by adding a resonance tube to the vocal tract[10]. Later, e.g. Titze et al.[11], Wade et al.[12] and Zañartu et al.[13] reported on occurrences of sudden pitch frequency jumps and other instabilities when the fundamental frequency of oscillation f_o was in the vicinity of the first vocal tract resonance frequency. These phenomena were also observed through numerical simulations by several authors[9,14–17]. Interactions with subglottal resonances might have a similar influence on the voice source waveform and the vocal fold vibrations as the vocal tract, as observed by Austin et al.[18] (using an excised larynx), Zhang et al.[19,20] (using vocal fold physical models and an excised larynx), or Lucero et al.[21] (using vocal fold physical and mathematical models).

Excised larynges allow obtaining deeper insight into the natural behavior of the voice source as their properties closely approximate those of the living larynges. Acoustic resonances inside the air-supplying tubes below the larynx (i.e., subglottal space), however, may influence the vibratory behavior and obscure the inherent vibratory properties of the larynx. To the best of our knowledge, no data exist revealing on how much the laryngeal vibratory behavior differs between anechoic and resonant subglottal conditions. In this paper, we therefore explore

Voice Research Lab, Department of Biophysics, Faculty of Science, Palacký University, 17. Listopadu 12, 771 46 Olomouc, Czechia. ✉email: jan.svec@upol.cz

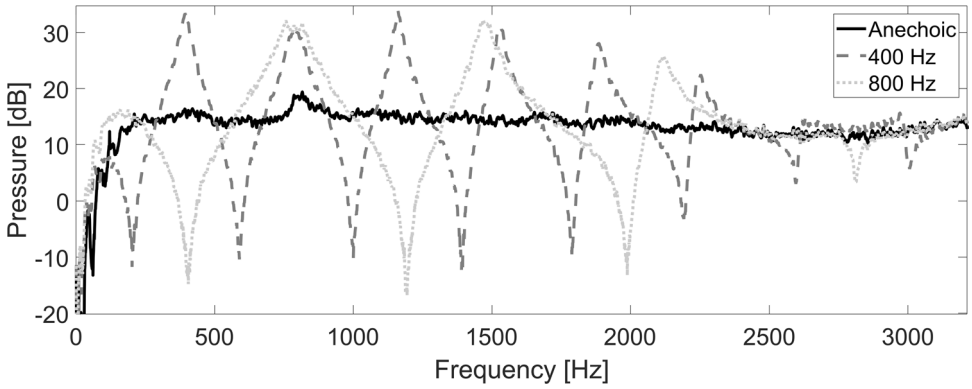


Figure 1. Frequency responses of the anechoic subglottal tract and of the resonant subglottal tract with the piston set to two different positions corresponding to $f_{R1} = 400$ and 800 Hz. Whereas the resonant tracts show multiple resonances and antiresonances (as expected for a tube with closed ends), the anechoic tract is practically free of these.

excised larynges with a newly developed anechoic (resonance-free) subglottal tract[22]. This setup eliminates the acoustic interactions with both the vocal and subglottal tracts and allows studying the vibration properties of the vocal folds in their inherent state. We measured the acoustic response of the newly developed anechoic tract and compared it to the acoustic response of an adjustable "resonant" subglottal tract, previously developed by Hampala et al.[23]. Finally, we used these subglottal tracts in excised larynx experiments where we measured the subglottal pressure waveforms and the radiated sound. The vocal fold vibrations were simultaneously monitored by the electroglottographic (EGG) signal, which is an approximate measure of the changes in the vocal fold contact area[24,25]. In these experiments, we compared the subglottal pressure waveforms and investigated the influence of the subglottal acoustics on the vocal fold vibrations and on the radiated sound.

Results

The acoustic responses of the subglottal tracts. To verify the functionality of the anechoic and resonant subglottal tracts we first measured their frequency responses. Figure 1 shows the responses for the anechoic subglottal tract and for the resonant subglottal tract set to two different resonance frequencies: $f_{R1} = 400$ and 800 Hz (only two resonance settings are presented here for simplicity). The resonant subglottal tract exhibits clear resonances and anti-resonances, as expected from the straight circular waveguide approximation: the lowest resonances and anti-resonances follow the patterns $f_{Rn} = nf_{R1}$ and $f_{ARn} = (2n—1)f_{AR1}$, where n is a positive integer. Importantly, the anechoic tract appears free of the acoustic resonances and has a response similar to one of an infinite, purely resistive waveguide. A small peak around 800 Hz was present in all the frequency responses, we therefore concluded that it was related to the damped resonance inside the air supply tubes.

Excised larynx experiments. *Pressure and EGG waveforms in the anechoic and resonant conditions: steady phonations.* To find out the effect of the anechoic and resonant subglottal tracts in excised larynx phonations we studied the subglottal pressure signal waveforms, detected with a pressure sensor just below the vocal folds. Those signals were obtained during steady phonations where the mean subglottal pressure had attained a saturation value, and the mean flow was set constant to about 400 mL s^{-1}. Figure 2 shows the subglottal pressure (a), EGG (b) and microphone (c) signal waveforms recorded while using the anechoic subglottal tract. The dashed vertical lines show the approximate instants of closure and opening of the vocal folds which were identified based on the subglottal pressure and EGG signal waveforms. As the anechoic subglottal tract is effectively equivalent to an infinite, purely resistive waveguide, it is expected that the pressure inside this waveguide is proportional to the flow[26]. Indeed, the subglottal pressure waveform resembled an inverted theoretical glottal flow voice source signal[27–32]: it was approximately constant during the closed phase, decreasing during the opening phase and increasing during the closing phase (Fig. 2a). To observe the similarity of the subglottal pressure waveform to the glottal flow signal, we performed inverse filtering analysis of the microphone signals. For this, we used the numerical integration feature offered by the Sopran software developed by Svante Granqvist[33], simulating the radiation impedance without any vocal tract. As expected, the resulting waveform was almost identical to the inverted subglottal pressure waveform obtained using the anechoic subglottal tract (Fig. 2d). The inverse filtered waveform was slightly more perturbed than the anechoic subglottal waveform, however.

In contrast to the anechoic tract, the subglottal pressure waveform appeared more complex when using the resonant subglottal tract. This is shown in Fig. 3 which compares the subglottal pressure waveforms (left, black solid lines) and the radiated sound waveforms (right, black solid lines) for the anechoic (Fig. 3a) and resonant subglottal tract set to six different resonance frequencies: $f_{R1} = 330$, 400, 500, 600, 700 and 800 Hz (Fig. 3b–g). These signals were scaled in time to show exactly three cycles for each signal. The glottal opening and closing instants were approximately synchronized using the corresponding EGG waveforms (grey dashed lines). In contrast to the anechoic subglottal pressure waveforms which showed very little fluctuations during the glottis closed phase, the presence of subglottal acoustic resonances introduced fluctuations of the subglottal pressure

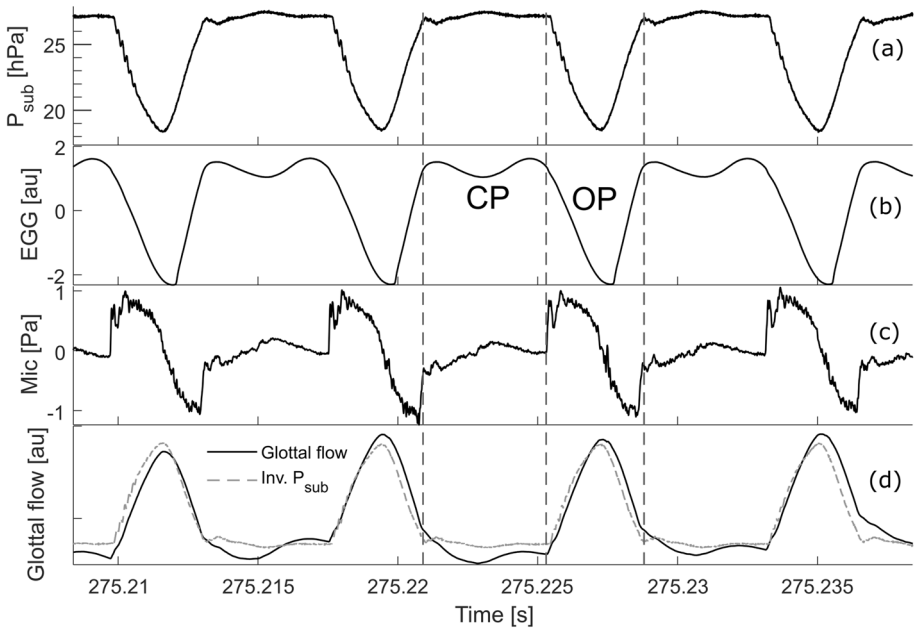


Figure 2. Signal waveforms for the excised larynx experiment in steady phonation, obtained using the anechoic subglottal tract. (**a**) Subglottal pressure; (**b**) EGG signal in arbitrary units [au]; (**c**) microphone signal (radiated sound) at 10 cm distance from the glottis; (**d**) glottal flow signal obtained by inverse filtering (black solid line) and inverted subglottal pressure (dashed gray line). The inverted subglottal pressure waveform in (**d**) was scaled to match the amplitude of the flow to observe the similarity of these waveforms. The closed phase (CP) and open phase (OP) were identified manually using both EGG and subglottal pressure signals.

during the closed phase similarly as observed previously in vivo[34–37] (see Fig. 3b–g, left). The frequency of the subglottal pressure fluctuations increased when f_{R1} increased, further indicating that they are caused by acoustic resonances in the subglottal tract. For the constant flow of 400 mL·s^{-1} the mean subglottal pressures (horizontal dashed black lines in Fig. 3, left) were about 25 hPa (i.e. c. 25 cm H_2O) in the anechoic conditions, whereas they were between 19 and 21 hPa in the resonant conditions. The radiated sound waveforms also showed changes with the different subglottal tract conditions (see Fig. 3, right), revealing that the subglottal tract influences also the radiated sound and thus indicating the presence of Level 1 interactions. These waveform changes were smaller compared to those of the subglottal pressure waveforms, however. For lower subglottal resonance frequencies, the EGG waveforms exhibited secondary peaks which approximately coincided with the peaks of the subglottal pressure waveforms (e.g. Fig. 3b,c). This suggests that the subglottal resonances influenced also the vocal fold vibrations, indicating Level 2 interactions. For higher subglottal resonance frequencies (Fig. 3d–g), the secondary EGG peaks did not occur and the EGG waveforms appeared nearly identical. Interestingly, the fundamental frequency of the vocal fold oscillations was found lowered in the resonant conditions (f_o around 106 Hz) compared to the anechoic ones (f_o around 126 Hz), even though the laryngeal settings were kept constant. The change of f_o between anechoic and resonant conditions suggests the presence of Level 2 interactions. Surprisingly, however, the f_o stayed around 106 Hz and did not change when the subglottal tract setting was changed among the six different resonance frequencies.

The spectra of the waveforms from Fig. 3 are presented in Fig. 4. Again, the subglottal spectra are on the left and the radiated spectra on the right. Clear harmonic components appeared in all the spectra at multiples of the fundamental frequency, revealing that the oscillations occurred at a steady pitch. The resonant subglottal spectra (Fig. 4b–g, left) exhibited a repetitive formant structure, reflecting the resonances and antiresonances observed in the frequency response of the resonant subglottal tracts (recall Fig. 1). The frequency of the first formant and the distance between the formants increased with the increasing resonance frequencies of the subglottal tract, as expected. The formant frequencies were, however, slightly lower than the resonance frequencies previously measured on the subglottal tracts without the larynx (indicated by dashed vertical lines in Fig. 4b–g). This difference can be explained by changes in the boundary conditions caused by the larynx. As expected, the envelope of the anechoic subglottal spectrum (Fig. 4a, left) was more uniform than in the resonant conditions and did not display a clear repetitive formant structure. The amplitude of the harmonic components decreased here with increasing frequency, although some fluctuations of the spectral envelope were also present.

The radiated sound spectra are shown in Fig. 4 on the right. In contrast to the resonant subglottal spectra, they did not display such a prominent formant structure in the resonant conditions. Nevertheless, slight modulation of the envelope of the radiated spectra occurred here too and the distances between the spectral envelope maxima appeared to increase with the increasing resonance frequency of the subglottal tract (Fig. 4b–g, right) similarly as in the subglottal spectra. This suggests that the subglottal resonances partially transferred to the radiated spectra. However, the subglottal resonance peaks were very broad and much less distinctive there. The

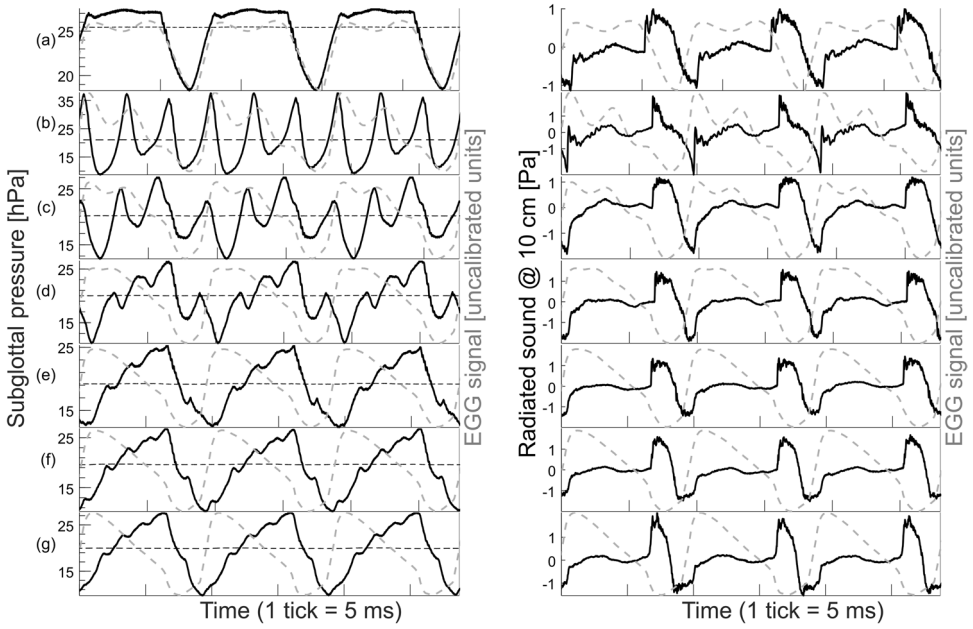


Figure 3. Subglottal pressure (left, solid black lines) and radiated sound (right, solid black lines) waveforms registered with an anechoic subglottal tract and with resonant subglottal tracts set to different resonance conditions: (**a**) anechoic subglottal tract, (**b**) $f_{R1} = 330$ Hz, (**c**) $f_{R1} = 400$ Hz, (**d**) $f_{R1} = 500$ Hz, (**e**) $f_{R1} = 600$ Hz, (**f**) $f_{R1} = 700$ Hz and (**g**) $f_{R1} = 800$ Hz. The horizontal dashed black lines on the left panel indicate the mean subglottal pressure in each case. The waveforms are individually scaled in time (3 cycles are shown). The instants of closure and opening were approximately synchronized using the corresponding EGG signal waveforms (dashed gray lines). Notice the drastic difference in the subglottal pressure waveforms between the anechoic (**a**) and resonant (**b**–**g**) cases. In (**b**–**g**), notice the increased frequency of the secondary subglottal oscillations related to the increasing resonance frequency of the subglottal tract. The radiated audio signals also show changes with the different subglottal tract conditions but these are smaller compared to those of the subglottal pressure and occur mainly during the glottal open phase.

radiated spectrum for the anechoic case (Fig. 4a, right) also showed some fluctuations in its envelope. It differed slightly from the radiated resonant spectra but these differences were, again, less distinct than the differences between the anechoic and resonant spectra in the subglottal space. As expected, the slope of the harmonic decay (the decrease in the peak amplitude for every doubling of the frequency) appeared smaller in the radiated sound spectra than in the subglottal sound spectra. This can be explained by the high-frequency amplification caused by the sound radiation into free air (see Eqs. (3) and (4) in the appendix).

Phonation changes due to interactions with the subglottal acoustics: flow sweeps. In order to find out whether there was an influence of the subglottal resonances on the vibrational properties of the vocal folds, we investigated the phonation threshold onset and offset pressures, the frequency of the vocal fold oscillations f_o, and the SPL of the subglottal and the radiated sounds. We analyzed the data from repeated flow sweeps, where the flow was slowly increased to about 550 mL·s^{-1} and slowly decreased back to zero. The experiments were done with an anechoic tract and with a resonant tract set to 500 Hz subglottal resonance frequency. In both the anechoic and resonant conditions, the offset pressures were generally smaller than the onset pressures (see Fig. 5). This agrees with the theoretical studies of vocal fold oscillation onset and offset[38,39] as well as with previous experimental observations on vocal fold mucosa[40] and excised larynges[41–43].

To find whether the subglottal resonance conditions had significant effect on the onset and offset phonation threshold values, we used multiple linear regression models (see the Supplementary material S1 for details on the statistics). The onset pressure values were not significantly different ($p = 0.97$) between the resonant and anechoic conditions. The offset pressure values were, however, approximately 11% lower (95% confidence interval 6–16%) in the resonant conditions than in the anechoic conditions, and this effect was statistically significant ($p = 0.0003$). These results suggest that the subglottal acoustics has little influence on the oscillation onsets, but has significant influence on the oscillation offsets.

Figure 6 shows the f_o values obtained from the phonations of the three larynges during the flow-sweep experiments. The f_o values were different for the different larynges. However the same phenomenon was visible in all the three larynges: the f_o values were consistently higher in the anechoic than in the resonant subglottal tract for the same subglottal pressures. This corroborates similar observation from the steady flow experiment where the f_o was also higher in the anechoic tract compared to all the resonant tracts. For clarity, only the values from the first and last of the flow sweeps are shown in Fig. 6, but these show the repeatability along the sweeps. Near the

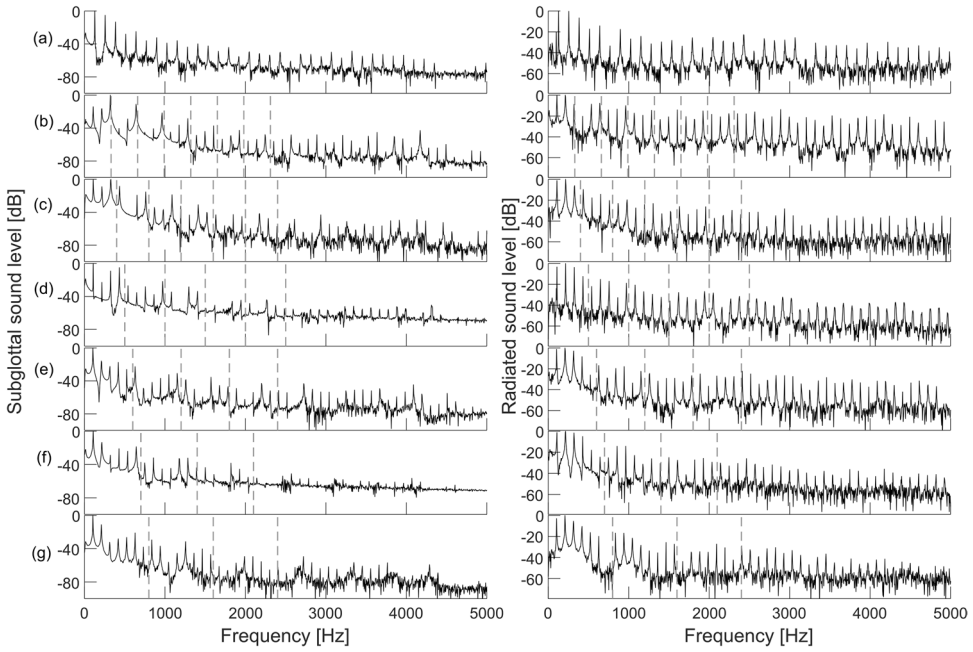


Figure 4. Spectra of the subglottal (left) and radiated (right) sounds, corresponding to the waveforms shown in Fig. 3. The subglottal settings were: (**a**) anechoic subglottal tract, (**b**) $f_{R1} = 330$ Hz, (**c**) $f_{R1} = 400$ Hz, (**d**) $f_{R1} = 500$ Hz, (**e**) $f_{R1} = 600$ Hz, (**f**) $f_{R1} = 700$ Hz and (**g**) $f_{R1} = 800$ Hz. The vertical dashed lines indicate the subglottal resonance frequencies up to 2.5 kHz, measured with no larynx attached. The spectra were produced at the frequency resolution of 4 Hz and were normalized by calculating the maximum value in each spectrum and setting it to 0 dB.

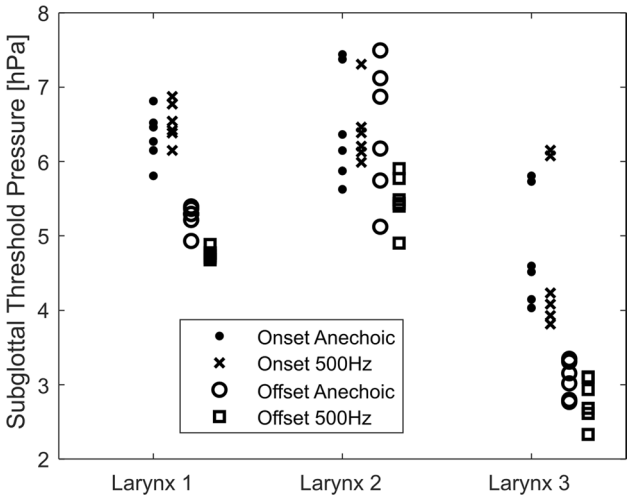


Figure 5. Oscillation threshold pressure for each larynx with an anechoic subglottal tract and resonant tract set to 500 Hz subglottal resonance. Each point corresponds to one flow sweep. Notice the onset thresholds are generally higher than the offset thresholds. The subglottal acoustics had little influence on the onset pressures (dots versus crosses), but the offset pressures were significantly lower for the resonant subglottal tract (circles versus squares). For the numerical values and the statistical results, see the Supplementary material S1 to this article.

phonation onsets and offsets, the larynges exhibited irregular vocal fold vibrations, therefore we did not include these parts of the signal in the f_o analysis. Only the parts with a stable f_o were kept for analysis.

For given mean subglottal pressures, the SPL of the subglottal sound was found to be higher when using the resonant tract, as shown in Fig. 7a–c. At the mean subglottal pressures above 8 hPa, the difference in SPL was about 6–8 dB for the first larynx, 2–3 dB for the second larynx and 4–5 dB for the third larynx. In both the anechoic and resonant cases the subglottal SPLs showed extremely high values reaching up to 150 dB re 20 µPa.

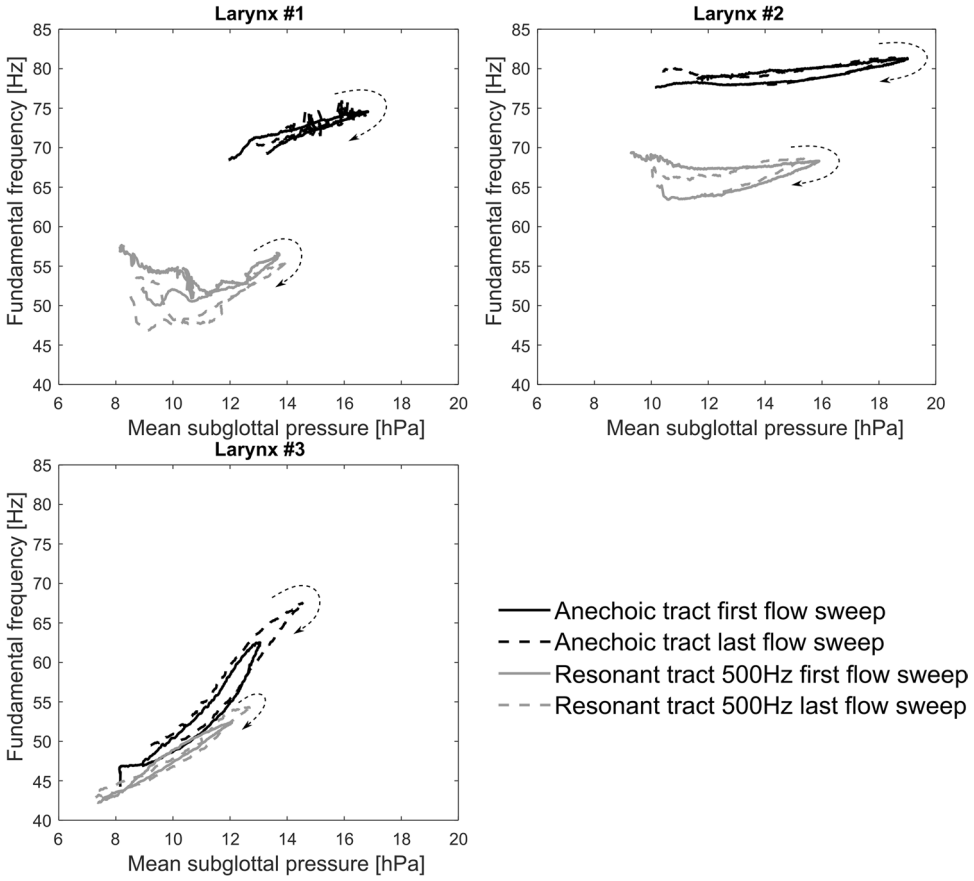


Figure 6. Analysis of f_o of the vocal fold oscillations with regards to the mean subglottal pressure. The results from the first (solid lines) and last (dashed lines) flow sweeps are showed, for each larynx. The black lines mark the data measured with the anechoic tract, whereas the gray lines mark the data measured with the resonant subglottal tract set to $f_{R1} = 500$ Hz. The dashed arrows indicate, for each flow sweep, the evolution of the measured values through time. In all the three larynges, the f_o values are lower for the resonant tract than for the anechoic one. The f_o differences vary among the three larynges, however.

Compared to the subglottal SPLs, the SPLs of the radiated sound at 10 cm distance were about 60 dB (!) lower reaching the maximum values of about 90 dB re 20 µPa. Similarly to the SPL of the subglottal sound, the radiated sound showed higher values when using the resonant tract, for the same subglottal pressures, as demonstrated in Fig. 7d–f. This increase was especially visible for mean subglottal pressure values above approximately 800 Pa. The increase in SPL, for identical mean subglottal pressures above 8 hPa, was about 2.5–3.5 dB for the first larynx, 2–4.5 dB for the second larynx and 1.5–2.5 dB for the third larynx.

Discussion

While subglottal resonances have been observed to influence voice and vocal fold vibrations[19], hardly any experimental data have been available documenting the voice and laryngeal behavior when the subglottal resonances are not present. Yet, removing the interactions with acoustic resonances is important for understanding the inherent vibratory properties of the voice source and of the vocal folds, and for validating computational models of voice production. To the best of our knowledge, Zhang et al.[44] did the first and so far the only study, which attempted to design and use an anechoic subglottal tract for voice generation. They used it for studying sound produced through an orifice simulating glottis with time-varying area. Anechoic terminations consisting of two connected perforated rubber hoses sealed with fiberglass were inserted into the air-supplying tube. Measurements of the frequency response revealed reduction of the subglottal resonances to some extent, although not fully. No excised larynx experiments using this tract were reported in the study.

In our experiments, we used a newly developed anechoic subglottal tract with a different design[22]. The measured frequency response (recall Fig. 1) revealed that the newly developed anechoic subglottal tract was able to cancel most of the acoustic resonances in the subglottal space. In this respect, the new design appears to be more effective than the anechoic subglottal tract pioneered by Zhang et al.[44]. Furthermore, the adjustable resonant subglottal tract allowed creating and changing the subglottal resonances to study their influence on the voice source and vocal fold behavior. The resonant subglottal tract allowed us to modify the lowest subglottal resonance frequencies from c. 330 to 800 Hz. This covers the range of subglottal resonance frequencies in humans, which are expected to be between 500 and 700 Hz[45–47]. The vocal fold fundamental frequency of 100–120 Hz

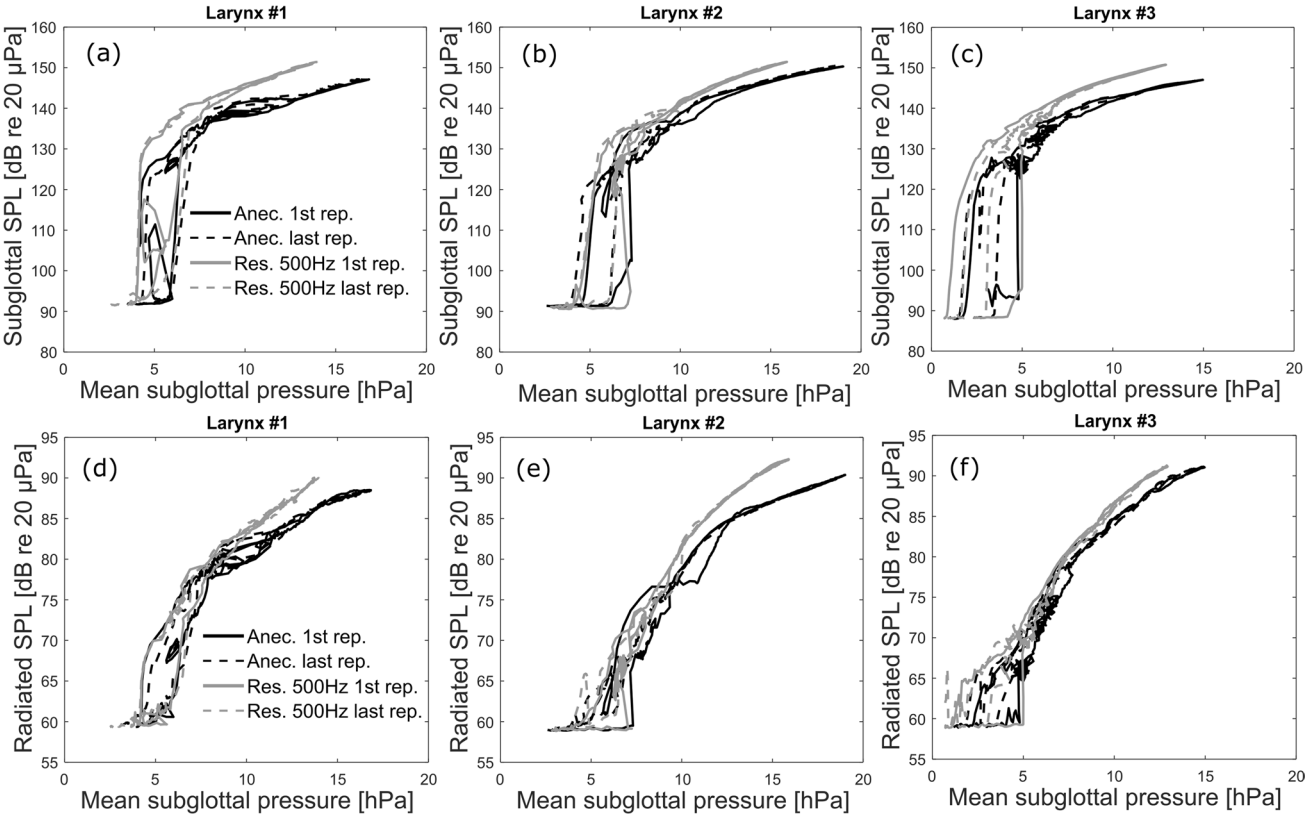


Figure 7. Analysis of the SPL (non-weighted) of the subglottal (**a**–**c**) and radiated (**d**–**f**) sounds at 10 cm distance with regards to the mean subglottal pressure. The black lines mark the data measured with the anechoic tract, whereas the gray lines mark the data measured with the resonant subglottal tract set to $f_{R1} = 500$ Hz. (**a**,**d**) First larynx, (**b**,**e**) second larynx, (**c**,**f**): third larynx. Notice that for a given subglottal pressure the SPLs are higher in the resonant than in the anechoic tract. For clarity, only the results from the first and last flow sweeps are shown, for each larynx. The solid lines correspond to the first flow sweep and the dashed lines to the last flow sweep for each larynx, illustrating the repeatability of the values.

measured in the red deer larynx during the steady phonation experiment corresponds to low-pitched phonations of male human subjects. The fundamental frequencies in the flow sweep experiment were lower, however, around 40–80 Hz, because larger red deer larynges were used.

To our knowledge, this study is the first to directly demonstrate the effect of absence and presence of subglottal resonances on phonation properties of excised larynges. Excised larynges are considered to be the most representative models of living vocal apparatus, therefore the observed effects can be expected to be similar to those observed in vivo. Our results clearly indicate that subglottal resonances influence both the radiated acoustic signals, as well as the vocal fold oscillations. Overall, for subglottal resonance frequencies that were much higher than the fundamental frequencies of vocal fold oscillations, their presence was found to slightly change the shape of the radiated acoustic waveforms (Fig. 3) and increase the radiated sound pressure levels by up to c. 4.5 dB (Fig. 7). The spectrum of the radiated sound showed some differences among the anechoic and resonant subglottal conditions. These differences were, however, considerably less apparent than those in the subglottal sound spectra which showed prominent formants in the resonant conditions (recall Fig. 4). The changes of the radiated pressure waveform, its SPL and spectrum belong to Level 1 interactions, but they can be influenced also by the Level 2 interactions as discussed below.

The presence of subglottal resonances was found also to change the fundamental frequency of the vocal fold oscillations (Fig. 6) and the threshold pressure for phonation offset (Fig. 5). These differences indicate the occurrence of Level 2 structure-acoustic interaction[9], as a change of subglottal acoustics induces a change in the vocal fold vibrations. Interestingly, no clear influence of subglottal resonances was found for the threshold pressure in phonation onset (Fig. 5). This suggests that the mechanisms for voice onset and offset should be seen differently from the perspective of subglottal interactions: subglottal resonances can be expected to be little excited in both the anechoic and resonant tracts before the phonation starts, thus having little influence on the phonation onset. At voice offset, however, the subglottal resonances are excited in a resonant tract, influencing the offset differently than the anechoic tract where no resonances occur.

The subglottal pressure oscillation amplitudes shown in Fig. 3 were around 800 Pa in the anechoic case and about twice as much in the resonant cases. These pressure fluctuations correspond to extremely high sound pressure levels around 140–150 dB re 20 µPa (non-weighted) which are indicated in Fig. 7. Such strong subglottal pressure oscillations appear to influence also the vocal fold tissue oscillations causing changes in the electroglottographic waveform (shown in Fig. 3), again indicating Level 2 interactions. Subglottal pressure oscillations of

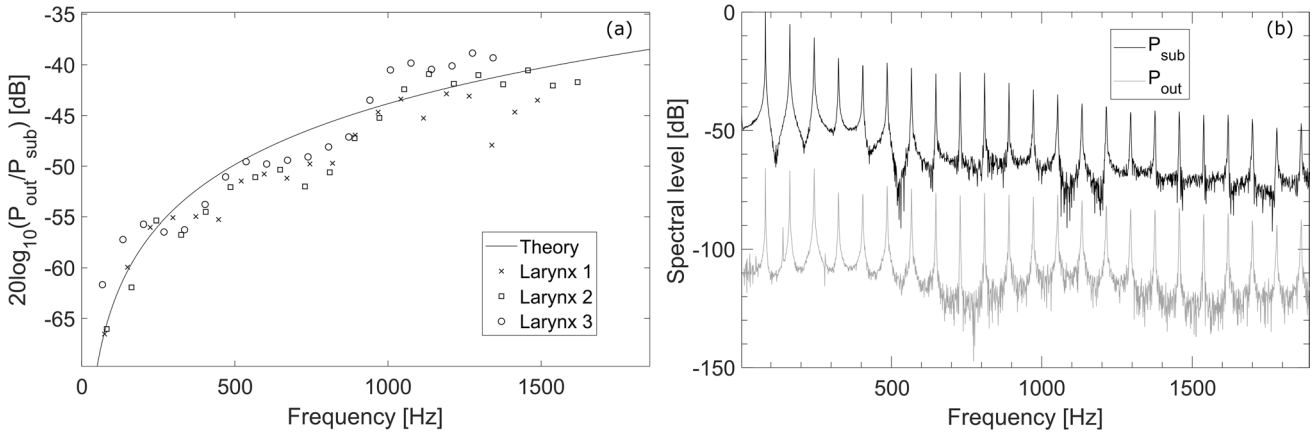


Figure 8. (**a**) Ratios of the first 20 spectral peak amplitudes between the radiated sound pressure and the subglottal pressure, expressed in dB, for the three larynges used during the flow sweeps experiment with anechoic subglottal tract. The solid line shows the expected ratio values according to the theory described in the appendix (value A_L from Eq. 5). (**b**) Example spectra of the subglottal (black line) and radiated (gray) sound pressures in the anechoic conditions. These spectra were taken from the second larynx, obtained from a one-second window at the maximum mean subglottal pressure during the first flow sweep, and normalized to the value of the first peak in the subglottal sound pressure spectrum.

similar magnitude were observed also in the in vivo human data[34,36,37,48,49]. The increase of the subglottal SPL in the resonant case (shown in Fig. 7a–c) is likely caused by the presence of subglottal acoustic resonances, which are boosting some harmonics of the source signal (see Fig. 4b–d, left). It is also possible that the vocal folds vibrated with more amplitude when using the resonant subglottal tract, which would also be a Level-2-interaction effect, but verifying this assumption would require accompanying laryngeal video recordings which were not available for these pilot experiments.

The radiated sound levels at 10 cm distance were about 60 dB lower than the subglottal sound levels. The highest radiated sound levels achieved here were around 90 dB (non-weighted) whereas the highest subglottal SPL were around 150 dB. The approximate 60 dB difference is a consequence of the conversion of the subglottal acoustic pressures to the acoustic volume flow which serves as the acoustic source radiating the sound to the surrounding space. The theoretical relationship between the subglottal and radiated sound pressures is derived in the Appendix. The radiation impedance for the glottal sound source is frequency dependent, therefore the low-frequency spectral components radiate less efficiently than the high-frequency components. Figure 8a compares the theoretical relationship between the subglottal and radiated pressures to the corresponding experimentally observed values in our excised larynx flow-sweep experiments. The values were obtained by comparing the spectral harmonic components of the subglottal and radiated sound (Fig. 8b). Even though the model relies on simplifying assumptions, the experimental and theoretical values follow the same trend and match reasonably well. The final SPL difference between the subglottal and radiated sound depends mainly on the dominant components of the sound spectrum which are around the lowest harmonic frequencies[50]. In our case these were around 50–200 Hz (Fig. 8b) where the decrease is, indeed, around 60 dB (Fig. 8a).

In the steady flow experiment shown in Fig. 3, we noticed that the mean subglottal pressures were higher for the anechoic than for the resonant conditions, in contrast to the oscillatory pressure amplitudes which showed an opposite tendency. Similarly, in the flow sweep experiments higher mean subglottal pressures were achieved in the anechoic conditions for the highest flows, as seen in Fig. 7. This suggests that glottal resistance is higher in anechoic than in the resonant conditions. Since the vocal fold settings were kept constant throughout the experiments, this is likely related to altered vocal fold oscillations (again indicating Level 2 interaction), possibly to their larger vibratory amplitudes in resonant conditions, but finding more about this effect would again require high-speed videolaryngoscopic data which were not available for these experiments. We plan to address this issue in future studies.

In an experiment with synthetic vocal folds and modifiable subglottal tract resonances, Zhang et al.[19] reported on a strong tendency of the vocal fold f_o to be driven towards the first, second, third or fourth subglottal resonance frequency. In our flow sweep experiments, we avoided such effects by keeping the subglottal resonances well above, i.e. 6–10 times higher than the vocal fold f_o, analogously as it is in the low-pitched phonations in humans. Interestingly, even under these conditions the presence of the subglottal resonances caused the vocal fold f_o to be decreased compared to anechoic conditions in all the cases investigated (Fig. 6). The amount of f_o change was different for the different larynges, however (Fig. 6). Once the resonant subglottal tract was attached, changing its resonance frequency between 330 and 800 Hz did not cause any significant change of f_o in the steady-flow-excised-larynx experiment (Fig. 3). The f_o remained to be around 106 Hz here. This suggests that no strong resonance tuning effects occurred between the subglottal resonances and the vocal fold oscillations. Nevertheless, the f_o differences between the anechoic and resonant subglottal conditions indicate that the mere presence of the resonances in the subglottal system can influence the vocal fold vibrations (Level 2 interaction). In future, it is desirable to confirm these effects also on human larynges.

The anechoic subglottal tract offers an interesting possibility to get direct information on the voice source waveform. As shown in Fig. 2, the vertically inverted subglottal pressure waveform matches well the inverse-filtered acoustic waveform. It shows the signal increasing and decreasing when the vocal folds open and close and being relatively constant during the glottis closure, as expected by the source-filter theory[51,52]. Compared to the subglottal-pressure-based waveform, the inverse-filtered waveform from the radiated sound shows more perturbations, suggesting the radiated sound is more polluted by surrounding noise and sound reflections from the structures around the larynx. The cleaner anechoic subglottal pressure waveform therefore appears to be advantageous for monitoring the voice source and can be explored in future studies.

Conclusion

The newly developed anechoic subglottal tract successfully removed its acoustic resonances to suppress their influence on vocal fold vibrations. When used in excised larynx experiments, the subglottal acoustic pressure waveform was similar to the inverted glottal flow source signal: almost constant during the closed phase, decreasing during the opening phase and increasing during the closing phase. In comparison, when using a resonant subglottal tract, the subglottal pressure exhibited fluctuations during the closed phase, which were related to the subglottal acoustic resonances, as expected. The subglottal resonances were found to influence both the radiated acoustic waveforms and radiated sound pressure levels, as well as the vocal fold oscillations. These provide direct evidence of the occurrence of both the Level 1 and Level 2 interactions of the voice source with subglottal pressure oscillations. The developed anechoic subglottal tract can be used to study the inherent properties of the voice source and vocal fold oscillations, free of acoustic interactions with adjacent cavities. The obtained data offer the basis for better understanding the inherent vibratory properties of the voice source, for studying the impact of structure-acoustic interactions on voice source, and for validation of simulation results obtained from computational models of voice production.

Methods

Design of the anechoic and resonant subglottal tracts and excised larynx experiments setup. The design of the subglottal tract was shortly described in a preliminary conference paper[22]. Here we provide its complete description. When designing the subglottal tract, we divided the subglottal spaces into a primary subglottal space and a secondary air-supply system. The primary subglottal space consisted of a 55 cm long straight cylindrical Plexiglas tube with the diameter of 24 mm. This tube formed an air space which could be changed from resonant to anechoic. Figure 9 shows simplified drawings and photographs of the subglottal tract in resonant and anechoic mode. In the resonant case, the tube was terminated by a piston the position of which could be changed to modify the resonance frequency of the subglottal space (Hampala et al.[23]). To change the resonant subglottal tract to the anechoic one, the piston was removed and the subglottal Plexiglas tube was extended by a plastic tube of the same diameter with the length of 330 cm terminated by a sound-absorbing pyramidal wedge. The wedge was approximately 200 cm long, and was made out of polyurethane foam (Molitan T-2337), with a density of about 23 $kg{\cdot}m^{-3}$. We closed this end of the tube with a plastic plug in order to prevent air leaks and pressure drops (see Fig. 9). Our subglottal tract is then seen by the voice source as a virtually infinite waveguide. The design of the sound-absorbing polyurethane wedge was inspired by the work of Sondhi (1975) who used similar approach to create an anechoic vocal tract for the purpose of inverse filtering[27].

The primary subglottal tract comprised a small side hole allowing the connection to the air-supply system using a small, rigid metal tube and then a long, flexible plastic tube. To prevent sound from entering or exiting the air supply system, the rigid metal tube was filled with polyurethane foam and a thinner thread of this foam (about 1 cm^2 cross-sectional area) was inserted in the flexible plastic tube (see Fig. 9). The thread was used to absorb the noise generated by the air pump and to cancel acoustic resonances inside the air-supply tubes as much as possible, in order to prevent any acoustic interaction with the larynx. The tubes had a 15 mm diameter.

An air pump (RESUN LP 100) was used to generate a continuous airflow through the system. The airflow was heated and humidified, before going through an artificial lung model (an air tank with a volume of 50 L with an inserted acoustically absorbing polyurethane foam) and to the subglottal tract via narrow air supply tubes. We used a mechanical flow meter placed just after the pump to adjust and monitor the amount of flow going out of the pump. The flow was also measured with a flow sensor placed after the flow meter. During the experiments, an electroglottograph (EGG) device (Glottal Enterprises EG2-PC) registered the vocal fold contact area, using two small electrodes screwed to the sides of the thyroid cartilage. We registered the subglottal acoustic pressure with a 2.4 mm diameter pressure transducer (Kulite XCQ093) placed at the inner side of the subglottal wall through a hole in the dorsal ridge of the cricoid cartilage. This pressure transducer is sensitive to frequencies between 0 and 150 kHz, allowing to measure both the DC pressure and the AC acoustic signals. The pressure transducer was inserted into a warmed metal tube to prevent moisture condensation around the transducer. A condenser microphone (MicW M416) placed approximately 10 cm above the glottis registered the radiated sound. All signals were sampled at 200 kHz using a DEWE-43 USB data acquisition system and recorded in the associated software Dewesoft X2.

Acoustic measurements. As the resonant subglottal tract has a shape of a simple straight circular waveguide, we expected it to exhibit harmonic resonance frequencies $f_{Rn} = nf_{R1}$, where f_{R1} is the first resonance frequency and f_{Rn} the n^{th} resonance frequency. Anti-resonances were also expected to be present, following the pattern $f_{ARn} = (2n—1)f_{AR1}$, where f_{AR1} is the first anti-resonance frequency and f_{ARn} the n^{th} anti-resonance frequency. We carried out pilot acoustic measurements to find the value of f_{R1} when the piston is at the lowest and highest positions, and to locate piston positions corresponding to 'round' resonance frequencies (400 Hz, 500 Hz, etc.). When doing this, we placed a small extension tube (about 3 cm long) at the upper open end of the

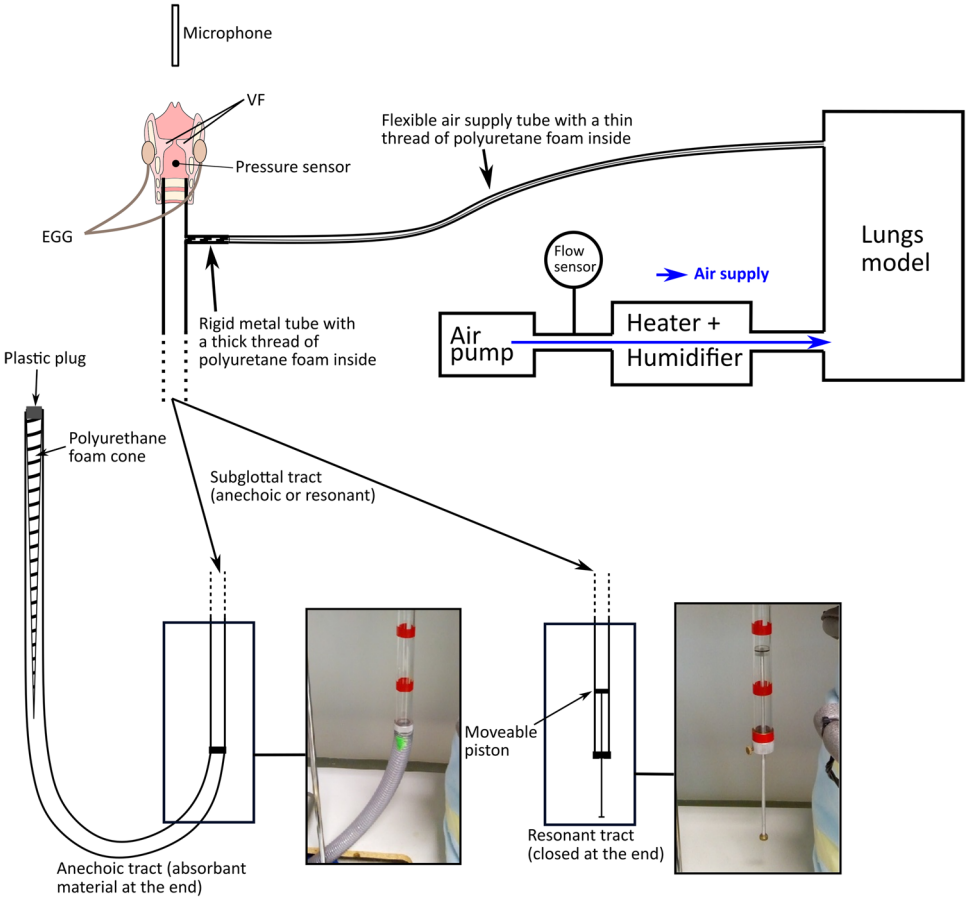


Figure 9. Excised larynx experimental setup.

subglottal system in order to approximate the space generally added by an excised larynx. To this extension tube, we attached a small electret microphone (AV-JEFE, TCM14) and a small loudspeaker (Ekulit LSF-23 M/N/G), using plasterine which enclosed the upper end to prevent air leaks. We used Audacity software (version 2.3.0)[53] to generate a ten-second linear chirp from 50 to 5000 Hz, to play it through the loudspeaker and to record the response of the subglottal tract captured by the microphone. Finally, we used the spectrum-plotting feature from Audacity to quickly find the approximate value of f_{R1}, by locating the first peak in the spectrum. We observed that f_{R1} was about 330 Hz when the piston was at the lowest position and about 800 Hz when the piston was at the highest position. We marked approximate positions of the piston corresponding to $f_{R1} = 400$, 500, 600 and 700 Hz with adhesive tape. These values were measured without a larynx specimen and might vary slightly when attaching one. The lengths of the subglottal tract corresponding to those values of f_{R1} were as follows: 46.5 cm (400 Hz), 37 cm (500 Hz), 31 cm (600 Hz) and 27 cm (700 Hz). As stated, the maximum length of the resonant subglottal tract was 55 cm (corresponding to $f_{R1} = 330$ Hz) and the minimum length was 25 cm (corresponding to $f_{R1} = 800$ Hz).

Consequently, we measured the acoustic responses of both subglottal tracts accurately using the same electret microphone and loudspeaker attached to the upper open end of the system. Again, we used plasterine to ensure the microphone and the loudspeaker were tightly fixed to the tract tube and to prevent air leaks. We used the following protocol to measure the acoustic response of both the anechoic and resonant subglottal tracts:

1. The loudspeaker played hundred impulses at the rate of one impulse per second. The impulse signal was manually generated by Audacity software.
2. The microphone registered the temporal response of all these impulses. The microphone signal was digitalized using a Focusrite Scarlett 2i2 2nd Gen USB audio interface and recorded by Audacity.
3. We segmented the microphone signal into one-second windows and averaged all the windows in the time domain to remove unwanted noise.
4. We performed Fast Fourier Transform (FFT) on the averaged temporal impulse response to get the frequency response of the system.

To compensate for the possible unevenness in the frequency responses of the microphone and loudspeaker, we measured the frequency response of the loudspeaker in free air with the same microphone, using the same

protocol. Then we divided the frequency responses of the subglottal tracts by the frequency response of the loudspeaker obtained in free air. We used Matlab custom-made scripts to perform all the numerical computations. The resonant subglottal tract was manually set to the previously determined piston positions corresponding to f_{R1} = 330, 400, 500, 600, 700 and 800 Hz.

Excised larynx experiments. *Preparation of the larynges.* During the excised larynx experiments we used red deer (*cervus elaphus*) larynges. These larynges were shown to behave similarly as human larynges[54,55]. The larynges were harvested from animals living wildly in forests, which were hunted by the Czech Army Forest Service during a regular hunting season, and they were treated in accordance with the standard ethical requirements of the Palacký University in Olomouc. After being harvested, the larynges were 'flash-frozen' using liquid nitrogen and kept in a freezer. Before the experiment, the larynges were put in a water bath heated to 30 °C until the larynx was completely defrosted. We prepared the larynges to expose the vocal folds, by removing the tissues above them: the epiglottis, the ventricular folds and part of the thyroid cartilage[56–58]. We attached the larynges at the open end of the subglottal tract and tightened them to prevent air leaks.

Experimental procedure. We used data from two separate experiments with red deer larynges. During these experiments, metal prongs were used to adduct the vocal folds and keep the laryngeal adjustment constant.

- In the first experiment (steady phonations), we adjusted the air flow to approximately 400 mL/s and waited until the subglottal pressure stabilized to a final value. After getting representative data we stopped the flow. We first performed the experiment with the anechoic subglottal tract attached to the system, then with the resonant subglottal tract set to six different values of f_{R1}: 330, 400, 500, 600, 700 and 800 Hz. To rule out the possibility of the long-term laryngeal tissue changes influencing the results, we then repeated the procedure once again with the anechoic subglottal tract and with the resonant subglottal tract set to the same values of f_{R1}, in order to verify the repeatability of the results.
- During the second experiment (flow sweeps), we first used the anechoic subglottal tract, and performed three flow sweeps, each executed in the following manner: the flow was slowly increased from 0 to approximately 550 mL/s, and after about five seconds we slowly decreased the flow back to 0 mL/s, for a total duration of about one minute per sweep. After that, we changed the subglottal tract to the resonant one (set to f_{R1} = 500 Hz), and repeated the three flow sweeps with the same flow values. Then, we switched back to the anechoic subglottal tract, and repeated the whole experiment once again to verify the repeatability of the results. Four larynges were used during the experiment, but one of them could not vibrate steadily, therefore we decided to discard it.

Data processing. The fundamental frequency f_o was estimated with the SWIPE' algorithm developed by Camacho et al.[59]. The onset and offset threshold pressures were measured by manually finding the oscillation onsets and offsets in the subglottal pressure waveform. We used a custom Matlab script to click on the waveform and get the onset and offset times. Afterwards, we averaged the subglottal pressure over a 50 ms window before the onset and after the offset. The radiated SPL was derived from the calibrated microphone signal at 10 cm from the vocal folds. We used the SPL calibration method 1A (using a calibrator and the microphone) described by Švec & Granqvist[50] together with the corresponding software package for Matlab[60]. The subglottal SPL was derived from the calibrated subglottal pressure signal, measured by the pressure sensor placed just below the vocal folds. The subglottal pressure was calibrated in cm H_2O using a U-shaped tube. The pressures were converted to hectopascals by multiplying it by the factor of 0.981. We applied a high-pass filter with a cutoff frequency of 10 Hz to the subglottal pressure and microphone signals to remove any DC offset, and then calculated both SPLs on the filtered signals using 'fast' time-weighting and no frequency weighting.

Inverse filtering. We performed inverse filtering analysis using the Sopran software developed by Svante Granqvist[33]. The signals were downsampled to 6 kHz to make the procedure easier. Through the software one can manually set up inverse poles and zeros to obtain a waveform as close to the theoretical source signal waveform as possible. Sopran also includes an option to select the type of the original signal: 'Sound pressure (mic)' or 'Flow signal (mask)'. When the first option is selected, Sopran adds a numerical integration step and a high-pass filter, for which the cutoff frequency is manually set by the user. As we used the microphone signal registering the radiated sound pressure, we accordingly selected the 'Sound pressure (mic)' option and used the cutoff frequency 20 Hz.

Statistics. Statistical tests were performed to find out significance of the differences between the anechoic and resonant subglottal conditions in the phonation onset and phonation offset pressures. Since we had onset and offset pressure data from two repetitions of three pressure sweeps for three larynges in both anechoic and resonant conditions, we utilized linear regression models with multiple categorical variables to take all these factors into account. We performed the statistical tests using Matlab built-in functions. The details on the statistical treatment of our data are provided in the Supplementary materials S1 to this article.

Data availability

The datasets generated and/or analyzed during the current study are available from the corresponding author on request.

Appendix

Theory of the relationship between the anechoic subglottal pressure, the glottal flow modulation and the radiated sound pressure. As mentioned in the methods section, the subglottal tract has the shape of a straight circular waveguide, and in the anechoic conditions it is assimilated to a semi-infinite waveguide. We can therefore assume that plane waves propagate in this waveguide in only one direction along the longitudinal axis of the waveguide. If we now consider the upwards direction (towards the glottis) along this axis to be the positive direction, then the subglottal acoustic waves propagate in the negative direction (away from the glottis). Below the glottis, the subglottal acoustic pressure p_s decreases when the glottis opens and increases when it closes, while the subglottal acoustic velocity u_s increases when the glottis opens and decreases when the glottis closes. This means p_s and u_s have an opposite phase. Furthermore, since in the anechoic conditions the waves propagate in only one direction, the amplitudes of p_s and u_s are proportionally related through the characteristic impedance $Z_c = \rho_0 c$[26], where ρ_0 is the mean density of air and c the velocity of sound in air. It is therefore possible to write the relation between both quantities:

$$u_s = -\frac{p_s}{\rho_0 c} = -\frac{p_s}{Z_c}. \tag{1}$$

For moist air with the temperature of 37 °C, it holds $Z_c \approx 400$ Pa·s·m^{-1}. If we express the subglottal volume flow U as the particle velocity u_s multiplied by the cross-sectional area of the anechoic subglottal tract S (in our case $S = 4.5$ mm^2), the volume flow can be expressed as:

$$U = -\frac{S}{Z_c} p_s. \tag{2}$$

Since S and Z_c are constants, the equation justifies that the acoustic pressure and the acoustic flow are proportional inside the anechoic subglottal tract and their waveform shapes are identical, but with reversed polarity.

The radiated acoustic pressure can be estimated from the oscillatory glottal volume flow. Because the glottal area is small compared to the wavelengths of the dominant frequencies of the voice spectrum, it is reasonable to approximate the oscillating glottis as a point source, generating the glottal volume flow Q. In the frequency domain, the glottal volume flow can be expressed as $Q = Q_0(\omega)e^{j\omega t}$, where $Q_0(\omega)$ is the amplitude, $j^2 = -1$, and ω is the angular frequency. Here Q only corresponds to the acoustic component of the glottal flow, disregarding the steady component which does not contribute to the sound. In this case, the radiated acoustic pressure p_r, in free air at the distance r from the glottis and at the angular frequency ω, can be estimated from the time derivative of the glottal volume flow by the theoretical relationship[61]:

$$p_r(r,\omega) = \frac{\rho_0}{4\pi r}\frac{\partial Q}{\partial t} = \frac{j\omega\rho_0 Q}{4\pi r}. \tag{3}$$

The complex notation is only useful here to know the phase difference (here 90 degrees) between the flow and the pressure. It is possible to express the volume flow U and the acoustic pressure p_s in the frequency domain through the Fourier series, as a sum of sinusoidal components of different amplitude, phase and frequency. If we assume that for every angular frequency ω, the subglottal volume flow $U(\omega)$ is equal to the glottal flow $Q(\omega)$, combination of Eqs. (2) and (3) provides an analytical relation between the anechoic subglottal pressure p_s and the radiated sound pressure p_r:

$$p_r(r,\omega) = -\frac{j\omega\rho_0 S}{4\pi r Z_c} p_s = -jA(r,\omega)p_s(\omega), \tag{4}$$

where $A(r,\omega) = \frac{\omega\rho_0 S}{4\pi r Z_c}$. To find the sound pressure level (SPL) difference between the radiated and subglottal sound in decibels, we can use the amplitudes of the radiated and subglottal pressure components ($p_{rA}(\omega)$ and $p_{sA}(\omega)$, respectively) and express their ratio logarithmically as:

$$A_L(\omega) = 20\log\frac{p_{rA}(\omega)}{p_{sA}(\omega)} = 20\log\frac{\omega\rho_0 S}{4\pi r Z_c}, \tag{5}$$

The theoretical dependence of A_L on frequency, for the distance $r = 10$ cm, was plotted by solid line in Fig. 8 for frequencies ranging from 50 to 2000 Hz.

Received: 16 July 2020; Accepted: 30 November 2020
Published online: 08 January 2021

References

1. Fant, G. M. *Acoustic theory of speech production*. (Mouton, 1960).
2. Van Den Berg, J. Myoelastic-aerodynamic theory of voice production. *J. Speech Lang. Hear. Res.* **1**, 227–244 (1958).
3. Titze, I. R. Comments on the myoelastic-aerodynamic theory of phonation. *J. Speech Lang. Hear. Res.* **23**, 495–510 (1980).
4. Titze, I. R. The physics of small-amplitude oscillation of the vocal folds. *J. Acoust. Soc. Am.* **83**, 1536–1552 (1988).
5. Titze, I. R. *The Myoelastic Aerodynamic Theory of Phonation*. (National Center for Voice and Speech, 2006).
6. Titze, I. R., Baken, R. J. & Herzel, H. Evidence of chaos in vocal fold vibration. in *Vocal Fold Physiology, Frontiers in Basic Science* 143–188 (edited by I. R. Titze (Singular Publishing Group, San Diego), 1993).
7. Baken, R. Between organization and chaos: a different view of the voice. In *Producing speech: Contemporary issues for Katherine Safford Harris* (ed Bell-Berti, F. & Raphael, L. J.) 233–245 (1995).

8. Herzel, H., Berry, D., Titze, I. & Steinecke, I. Nonlinear dynamics of the voice: Signal analysis and biomechanical modeling. *Chaos Interdiscip. J. Nonlinear Sci.* **5**, 30–34 (1995).
9. Titze, I. R. Nonlinear source-filter coupling in phonation: theory. *J. Acoust. Soc. Am.* **123**, 2733–2749. https://doi.org/10.1121/1.2832337 (2008).
10. Weiss, D. Zur Frage der Registerbruchstellen. Die Wirkung vorgeschalteter Resonanzröhren auf die Stimme. [On the question of register breaks. The effect of upstream resonance tubes on the voice]. *Z Hals-, Nas-u Ohrenheilk* **70**, 353–358 (1932).
11. Titze, I. R., Riede, T. & Popolo, P. Nonlinear source–filter coupling in phonation: Vocal exercises. *J. Acoust. Soc. Am.* **123**, 1902–1915 (2008).
12. Wade, L., Hanna, N., Smith, J. & Wolfe, J. The role of vocal tract and subglottal resonances in producing vocal instabilities. *J. Acoust. Soc. Am.* **141**, 1546. https://doi.org/10.1121/1.4976954 (2017).
13. Zañartu, M., Mehta, D. D., Ho, J. C., Wodicka, G. R. & Hillman, R. E. Observation and analysis of in vivo vocal fold tissue instabilities produced by nonlinear source-filter coupling: a case study. *J. Acoust. Soc. Am.* **129**, 326–339. https://doi.org/10.1121/1.3514536 (2011).
14. Tokuda, I. T. Non-linear dynamics in mammalian voice production. *Anthropol. Sci.* **126**, 35–41 (2018).
15. Murtola, T., Aalto, A., Malinen, J., Aalto, D. & Vainio, M. Modal locking between vocal fold oscillations and vocal tract acoustics. *Acta Acust. united Ac.* **104**, 323–337 (2018).
16. Mergell, P. & Herzel, H. Modelling biphonation—the role of the vocal tract. *Speech Commun.* **22**, 141–154 (1997).
17. Hatzikirou, H., Fitch, W. & Herzel, H. Voice instabilities due to source-tract interactions. *Acta Acust. united Ac.* **92**, 468–475 (2006).
18. Austin, S. F. & Titze, I. R. The effect of subglottal resonance upon vocal fold vibration. *J. Voice* **11**, 391–402 (1997).
19. Zhang, Z., Neubauer, J. & Berry, D. A. The influence of subglottal acoustics on laboratory models of phonation. *J. Acoust. Soc. Am.* **120**, 1558–1569 (2006).
20. Zhang, Z., Neubauer, J. & Berry, D. A. Influence of vocal fold stiffness and acoustic loading on flow-induced vibration of a single-layer vocal fold model. *J. Sound Vib.* **322**, 299–313 (2009).
21. Lucero, J. C., Lourenço, K. G., Hermant, N., Van Hirtum, A. & Pelorson, X. Effect of source–tract acoustical coupling on the oscillation onset of the vocal folds. *J. Acoust. Soc. Am.* **132**, 403–411 (2012).
22. Lehoux, H., Hampala, V. & Švec, J. G. Development and Use of an Anechoic Subglottal Tract for Excised Larynx Experiments. in *Models and Analysis of Vocal Emissions for Biomedical Applications - 11th International Workshop, MAVEBA 2019.* (ed Claudia Manfredi) 209–212 (Firenze University Press, 2019).
23. Hampala, V., Švec, J., Schovánek, D. & Mandát, D. Utility Model No. 25585: Subglottal tract model (In Czech). Czech republic patent (2013).
24. Baken, R. J. Electroglottography. *J. Voice* **6**, 98–110 (1992).
25. Hampala, V., Garcia, M., Švec, J. G., Scherer, R. C. & Herbst, C. T. Relationship between the electroglottographic signal and vocal fold contact area. *J. Voice* **30**, 161–171 (2016).
26. Kinsler, L. E., Frey, A. R., Coppens, A. B. & Sanders, J. V. The Acoustic Wave Equation and Simple Solutions. in *Fundamentals of Acoustics* 113–148 (John Wiley & Sons, Inc. New York, 2000).
27. Sondhi, M. M. Measurement of the glottal waveform. *J. Acoust. Soc. Am.* **57**, 228–232 (1975).
28. Miller, R. L. Nature of the vocal cord wave. *J. Acoust. Soc. Am.* **31**, 667–677 (1959).
29. Murtola, T., Alku, P., Malinen, J. & Geneid, A. Parameterization of a computational physical model for glottal flow using inverse filtering and high-speed videoendoscopy. *Speech Commun.* **96**, 67–80 (2018).
30. Rothenberg, M. A new inverse-filtering technique for deriving the glottal air flow waveform during voicing. *J. Acoust. Soc. Am.* **53**, 1632–1645 (1973).
31. Sundberg, J. Flow glottogram and subglottal pressure relationship in singers and untrained voices. *J. Voice* **32**, 23–31 (2018).
32. Wolfe, J., Chu, D., Chen, J.-M. & Smith, J. An experimentally measured source-filter model: glottal flow, vocal tract gain and output sound from a physical model. *Acoust. Aust.* **44**, 187–191. https://doi.org/10.1007/s40857-016-0046-7 (2016).
33. Granqvist, S. Sopran sound editor, available from http://www.tolvan.com/index.php?page=/main/home.php (last accessed 02.06.2020).
34. Cranen, B. & Boves, L. Pressure measurements during speech production using semiconductor miniature pressure transducers: Impact on models for speech production. *J. Acoust. Soc. Am.* **77**, 1543–1551 (1985).
35. Miller, D. & Schutte, H. Characteristic patterns of sub-and supraglottal pressure variations within the glottal cycle. In *Transcr. XIIIth Symp. Care Prof. Voice.* 70–75 (1985).
36. Schutte, H. & Miller, D. Resonanzspiele der Gesangsstimme in ihren Beziehungen zu supra-und subglottalen Druckverlaufen: Konsequenzen für die Stimmbildungstheorie. [Play of Resonances in the Singing Voice in the Supra- and Subglottal Pressure Changes: Consequences for the Theory of Voice Production]. *Folia Phoniatr. Logo.* **40**, 65–73 (1988).
37. Sundberg, J., Scherer, R., Hess, M., Müller, F. & Granqvist, S. Subglottal pressure oscillations accompanying phonation. *J. Voice* **27**, 411–421 (2013).
38. Lucero, J. C. Subcritical hopf bifurcation at phonation onset. *J. Sound Vib.* **218**, 344–349 (1998).
39. Lucero, J. C. Oscillation hysteresis in a two-mass model of the vocal folds. *J. Sound Vib.* **282**, 1247–1254 (2005).
40. Titze, I. R., Schmidt, S. S. & Titze, M. R. Phonation threshold pressure in a physical model of the vocal fold mucosa. *J. Acoust. Soc. Am.* **97**, 3080–3084 (1995).
41. Berry, D. A., Herzel, H., Titze, I. R. & Story, B. H. Bifurcations in excised larynx experiments. *J. Voice* **10**, 129–138 (1996).
42. Mau, T., Muhlestein, J., Callahan, S., Weinheimer, K. T. & Chan, R. W. Phonation threshold pressure and flow in excised human larynges. *Laryngoscope* **121**, 1743–1751 (2011).
43. Regner, M. F., Tao, C., Zhuang, P. & Jiang, J. J. Onset and offset phonation threshold flow in excised canine larynges. *Laryngoscope* **118**, 1313–1317 (2008).
44. Zhang, Z., Mongeau, L. & Frankel, S. H. Experimental verification of the quasi-steady approximation for aerodynamic sound generation by pulsating jets in tubes. *J. Acoust. Soc. Am.* **112**, 1652–1663 (2002).
45. Arsikere, H., Leung, G. K., Lulich, S. M. & Alwan, A. Automatic estimation of the first three subglottal resonances from adults' speech signals with application to speaker height estimation. *Speech Commun.* **55**, 51–70 (2013).
46. Cranen, B. & Boves, L. On subglottal formant analysis. *J. Acoust. Soc. Am.* **81**, 734–746 (1987).
47. Ishizaka, K., Matsudaira, M. & Kaneko, T. Input acoustic-impedance measurement of the subglottal system. *J. Acoust. Soc. Am.* **60**, 190–197 (1976).
48. Miller, D. G. & Schutte, H. Effects of downstream occlusions on pressures near the glottis in singing. in *Vocal fold physiology: Acoustic, perceptual, and physiological aspects of voice mechanism* 91–98 (1991).
49. Schutte, H. K. & Miller, D. G. The effect of F0/F1 coincidence in soprano high notes on pressure at the glottis. *J. Phon.* **14**, 385–392 (1986).
50. Svec, J. G. & Granqvist, S. Tutorial and guidelines on measurement of sound pressure level in voice and speech. *J. Speech Lang. Hear. Res.* **61**, 441–461. https://doi.org/10.1044/2017_JSLHR-S-17-0095 (2018).
51. Alku, P. Glottal inverse filtering analysis of human voice production: A review of estimation and parameterization methods of the glottal excitation and their applications. *Sadhana* **36**, 623–650 (2011).
52. Fant, G., Liljencrants, J. & Lin, Q.-G. A four-parameter model of glottal flow. *STL-QPSR* **4**, 1–13 (1985).
53. The Audacity Team. Audacity: Free Audio Editor and Recorder. At https://www.audacityteam.org/ (2020).

54. Herbst, C. T. Glottal efficiency of periodic and irregular in vitro red deer voice production. *Acta Acust. United Ac.* **100**, 724–733 (2014).
55. Herbst, C. T., Hampala, V., Garcia, M., Hofer, R. & Svec, J. G. Hemi-laryngeal setup for studying vocal fold vibration in three dimensions. *J. Vis. Exp.* https://doi.org/10.3791/55303 (2017).
56. Durham, P. L., Scherer, R., Druker, D. & Titze, I. *Development of Excised Larynx Procedures for Studying Mechanisms of Phonation*. (Technical report: Voice Acoustics and Biomechanics Laboratory, Department of Speech Pathology and Audiology, The University of Iowa, IA, USA., 1987).
57. Titze, I. R. Experiments with excised larynges. In *The myoelastic aerodynamic theory of phonation* 1–62 (National Center for Voice and Speech, 2006).
58. van den Berg, J. & Tan, T. S. Results of experiments with human larynxes. *Pract. Oto-Rhino-Laryng.* **21**, 425–450. https://doi.org/10.1159/000274240 (1959).
59. Camacho, A. & Harris, J. G. A sawtooth waveform inspired pitch estimator for speech and music. *J. Acoust. Soc. Am.* **124**, 1638–1652 (2008).
60. Granqvist, S. & Švec, J. G. CalibrateVoiceSPL. Matlab scripts available from http://www.mathworks.com/matlabcentral/fileexchange/64231-calibratevoicespl (last accessed 02.06.2020) (2017).
61. Kinsler, L. E., Frey, A. R., Coppens, A. B. & Sanders, J. V. Radiation and Reception of Acoustic Waves. in *Fundamentals of Acoustics* 171–209 (John Wiley & Sons, Inc. New York, 2000).

Acknowledgements

This work was supported by the Czech Science Foundation (GA CR) project no. 19-04477S. A preliminary version of this work was presented at the 11th International Workshop Models and Analysis of Vocal Emissions for Biomedical Applications (MAVEBA), in Florence, Italy on December 17-19, 2019. The authors appreciate the help of O. Vencálek, Ph.D. from the Department of Mathematical Analysis and Applications of Mathematics at the Faculty of Science of the Palacký University in Olomouc with the statistical analysis of the measured data, and the help of Ing. P. Liška from the Czech Army Forest Service with acquiring the excised deer larynges.

Author contributions

H.L.: formal analysis, investigation, data curation, writing—original draft, visualization. V.H.: methodology, subglottal tract model development, investigation. J.G.Š.: conceptualization, methodology, resources, writing—review and editing, supervision, funding acquisition.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-79265-3.

Correspondence and requests for materials should be addressed to J.G.Š.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplement B: Paper II

Frequency jumps in excised larynges in anechoic conditions: A pilot study.

H. LEHOUX, C. T. HERBST, M. DOBIÁŠ, J. G. ŠVEC

Journal of Sound and Vibration, **551**, 117607 (2023)

doi: 10.1016/j.jsv.2023.117607

Journal of Sound and Vibration 551 (2023) 117607

Contents lists available at ScienceDirect

Journal of Sound and Vibration

journal homepage: www.elsevier.com/locate/jsvi

Frequency jumps in excised larynges in anechoic conditions: A pilot study

Hugo Lehoux [a], Christian T. Herbst [b,c], Martin Dobiáš [d,e], Jan G. Švec [a,*]

[a] Department of Experimental Physics, Faculty of Science, Palacký University in Olomouc, Czechia
[b] Department of Vocal Studies, Mozarteum University, Salzburg, Austria
[c] Janette Ogg Voice Research Center, Shenandoah Conservatory, Winchester, VA, United States of America
[d] Department of Forensic Medicine and Medical Law, Faculty of Medicine and Dentistry, Palacký University, Olomouc, Czechia
[e] Department of Forensic Medicine and Medical Law, University Hospital Olomouc, Olomouc, Czechia

ARTICLE INFO

Keywords:
Excised larynx
Voice
Voice registers
Frequency jumps
Bifurcation

ABSTRACT

Sudden fundamental frequency jumps between chest and falsetto registers are among the least understood phenomena occurring in human voice. Such jumps are recognized as bifurcation events and have been assumed to result from nonlinear-dynamic properties of the vocal folds interacting with acoustic resonances of the vocal tract and of the subglottal tract (ST). Here, we explored an anechoic (resonance-free) ST and investigated these frequency jumps in five excised human larynges, without a supraglottal vocal tract, under two conditions: (a) anechoic and (b) subglottally resonant. When smoothly elongating the vocal folds, we observed consistent jumps in anechoic conditions, proving that subglottal and supraglottal vocal tract resonances are not necessary for the jumps to arise. The presence of a resonant ST did not result in more numerous jumps compared to anechoic conditions, indicating that the inherent nonlinear-dynamic properties of the larynges were the primary cause for the jumps. Nevertheless, the resonant ST slightly altered the initial and terminating frequencies of the jumps, suggesting that the role of the interaction of the vocal fold oscillations with subglottal acoustics should not be neglected. These experimental findings should be considered when validating mathematical models simulating the nonlinear-dynamic behavior of the human vocal apparatus.

1. Introduction

Breaks between chest and falsetto registers accompanied by sudden fundamental frequency (f_o) jumps occur typically in adolescent male voices but can also be found in voice disorders or some singing styles [1,2]. Such jumps can occur abruptly without any abrupt changes in phonatory settings [1,2]. They have been described as bifurcation events that emerge from inherent nonlinear dynamic properties of the phonatory apparatus [1], but they have to date not been explained in sufficient detail. Experimental studies have identified multiple factors which stimulate abrupt f_o jumps, such as the interactions between (a) vocal fold (VF) vibrations and the acoustic resonances of the supraglottal vocal tract [3,4], (b) VF vibrations and the acoustic resonances of the subglottal tract (ST) [5], and (c) the vibrations of the left and right VFs [6]. Theoretical and modeling studies predicted that sudden frequency jumps might occur without any influence of subglottal or supraglottal vocal tract resonances [7,8], but these predictions have never been verified

* Corresponding author.
E-mail address: jan.svec@upol.cz (J.G. Švec).

https://doi.org/10.1016/j.jsv.2023.117607
Received 9 November 2022; Received in revised form 9 February 2023; Accepted 10 February 2023
Available online 12 February 2023

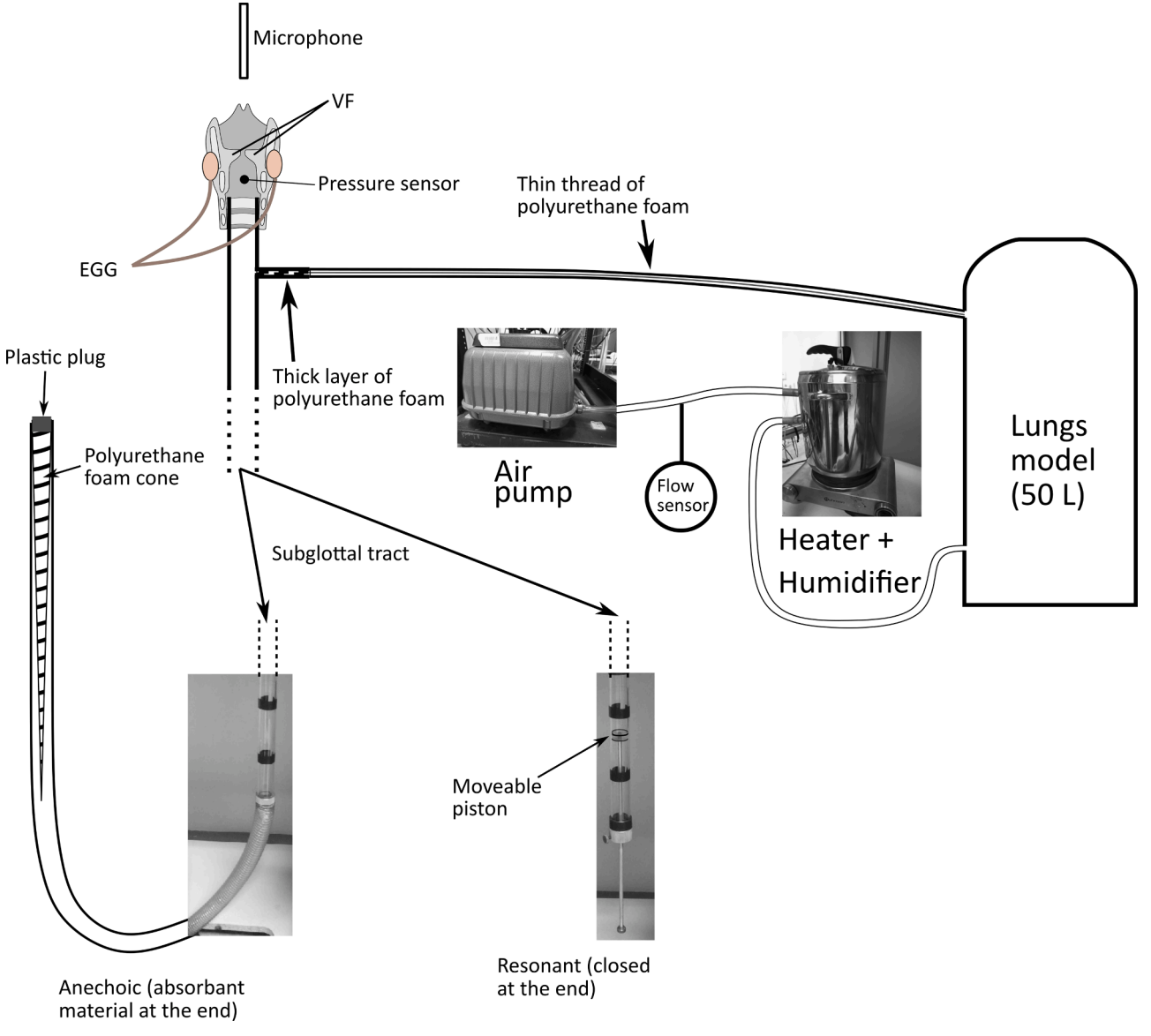


Fig. 1. Experimental setup for the excised larynx experiments.

empirically with real larynges.

In order to empirically investigate the inherent laryngeal vibratory properties, we have developed an excised larynx setup with an anechoic (i.e., resonance-free) ST [9]. This setup allows eliminating interactions of the subglottal acoustic resonances with the VFs. In this pilot study, we investigated the occurrence of abrupt f_o jumps in excised human larynges in two conditions: (a) fully **anechoic** and (b) subglottally **resonant**, without any supraglottal vocal tract. The goal of this study was to determine the potential influence of subglottal resonances on the f_o jumps when smoothly elongating the VFs. In particular, we wanted to find out whether the abrupt jumps occur also when subglottal resonances are not present.

2. Methods

We used 5 excised larynges from deceased male humans (aged 55, 64, 69, 72, and 77 years), for whom a medical autopsy was ordered for forensic purposes. The bodies had been placed in a standard refrigerated facility within a short time interval after death and did not show any signs of postmortem autolytic changes at autopsy. Autopsies were performed within 24–48 h after death. After the required standard diagnostic procedures of the medical autopsy had been fulfilled, the larynges were secured for further investigations in accordance with the ethical rules established by Czech legislation [10] and used for this study. Secured specimens were flash frozen with liquid nitrogen and stored at −80 °C. Each larynx was placed in a refrigerator at about 4 °C the night before the experiment to facilitate slow thawing. On the day of the experiment, each larynx was placed in a water bath at 30 °C until completely thawed. A visual inspection was performed to ensure that only larynges without abnormalities were used.

We prepared the larynges to expose the VFs by removing the superficial tissues above them, following previously established guidelines [11,12]. The upper part of the thyroid cartilage was cut, but the anterior notch was kept in order to facilitate the elongation of the VFs. Metal prongs were used to adduct the arytenoids (as described in [12]).

The experimental setup is presented in Fig. 1. A detailed description of this setup can be found in Lehoux et al. [9]; here we provide its brief description. The ST consists of a plexiglass tube of 55 cm length, which can be set to either a **resonant** (RES) or **anechoic** (AE) condition by modifying its termination. For the anechoic condition, the base plexiglass tube was extended by a long plastic tube terminated by a pyramidal wedge made from polyurethane that absorbed the sound waves and prevented acoustic resonances. For the resonant condition, a rigid moveable piston terminated the plexiglass tube.

For the present experiment, we set the first resonance frequency f_{R1} to 500 Hz, so that it approximated the values observed *in vivo* [13]. The frequency response of the subglottal tract was measured using a small loudspeaker and a small electret microphone attached to the glottal end of the tract, fixed with plasticine to prevent air leaks. We obtained the final response by dividing it by the response of the loudspeaker, measured in free space. More details on the procedure can be found in [9] (Methods section, pp. 9–11). The frequency responses in the anechoic and resonant conditions are shown in Fig. 2.

For the excised larynx experiments, we used a condenser microphone (MicW M416, Beijing Shengwang Acoustics Technology Co., Beijing, China) to register the radiated sound, calibrated at 94 dB re 20 μPa (1 kHz tone). The airflow was monitored and controlled by a F300L mechanical flow meter (GM Instruments, Irvine, United Kingdom) and registered with an SDP-1000 L-05 flow sensor (Sensirion AG, Stäfa, Switzerland). The subglottal pressure was measured using two pressure sensors: a Kulite XCQ-093 (Kulite Semiconductor Products, Inc., Leonia, NJ, USA) and a Keller 41X (Keller Group plc, London, UK). Both pressure sensors were calibrated in cm H_2O using a U-tube. Electroglottography (EGG) was used to monitor the VF contact area (not analyzed here for brevity). All signals were digitized using a DEWE-43 USB data acquisition system (DEWESOFT d.o.o., Trbovlje, Slovenia) at 48 kHz and recorded by the associated software Dewesoft X2 (DEWESOFT d.o.o., Trbovlje, Slovenia).

We performed repetitive frequency sweeps by slowly elongating and shortening the VFs. This was accomplished by manually pulling and pushing the thyroid notch, respectively. The following four tasks were executed:

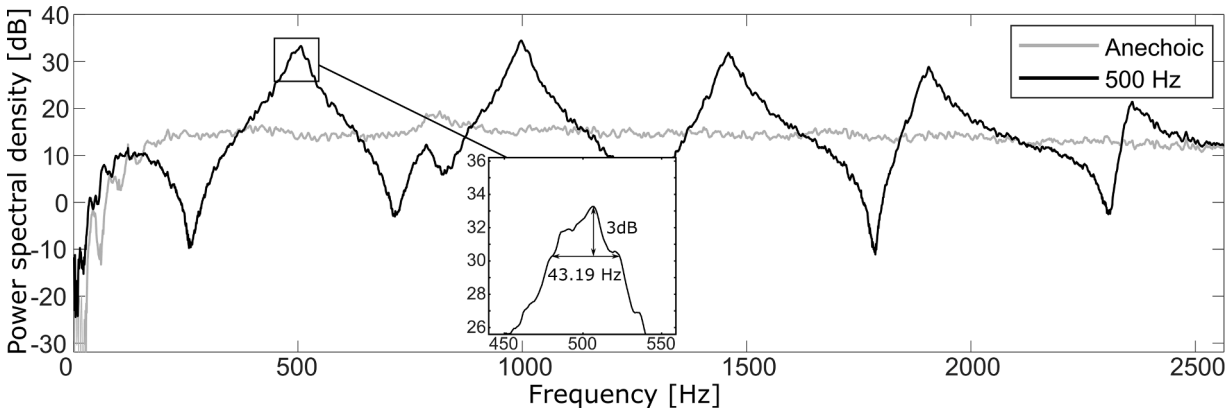


Fig. 2. Frequency response of the subglottal tract (ST) in anechoic (gray line) and resonant (black line) setup. The bandwidth of the first subglottal resonance is also shown.

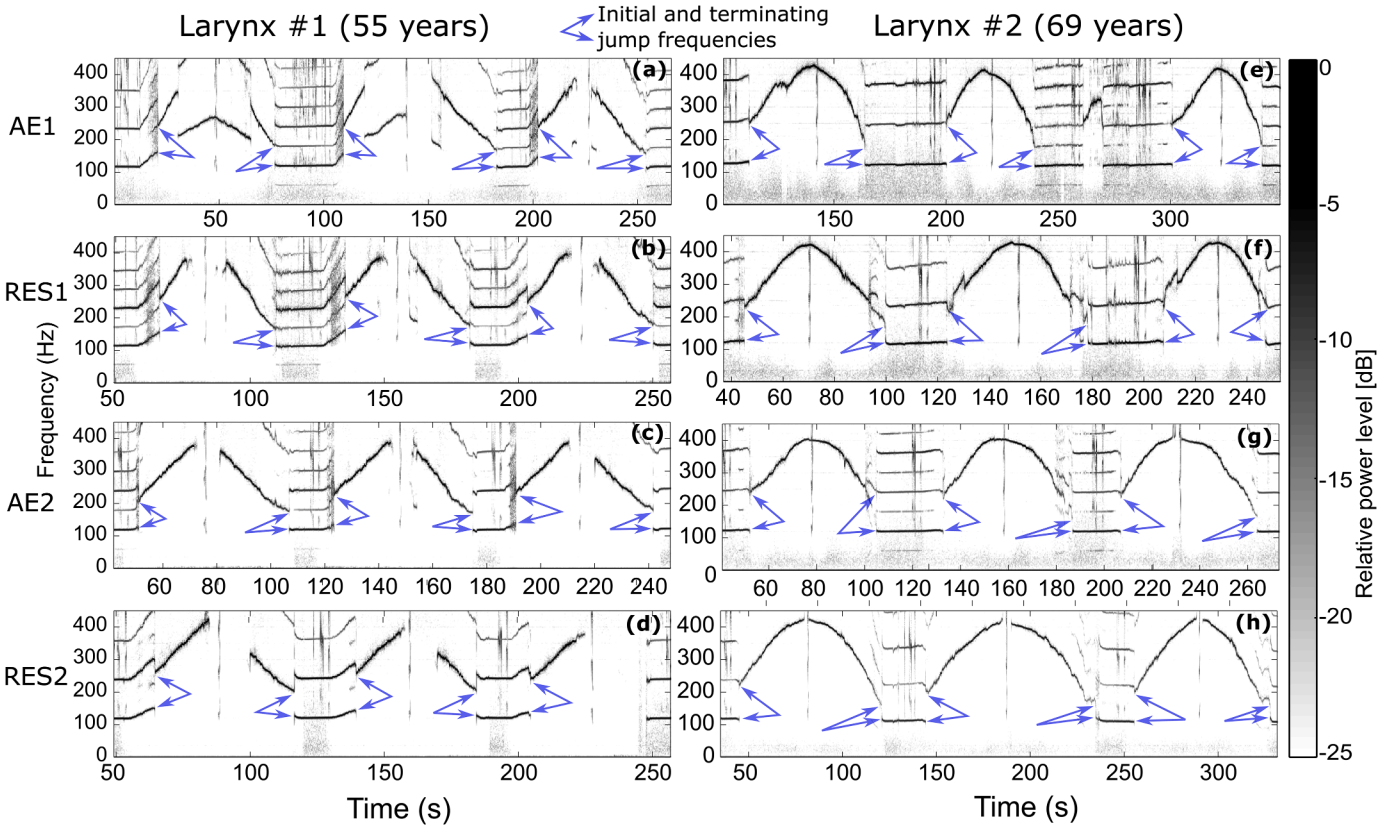


Fig. 3. Spectrograms derived from the microphone signal for each recording. The blue arrows indicate the initial and terminating frequencies of the investigated upward and downward f_o jumps. AE1 – 1st anechoic series: three elongation-shortening sweeps with the anechoic ST; RES1 – 1st resonant series: three elongation-shortening sweeps with the resonant ST set to $f_{R1} = 500$ Hz; AE2 – 2nd anechoic series: repetition of AE1; RES2 – 2nd resonant series: repetition of RES1.

(1) **AE1** – 1st anechoic series: three elongation-shortening sweeps with the anechoic ST
(2) **RES1** – 1st resonant series: three elongation-shortening sweeps with the resonant ST
(3) **AE2** – 2nd anechoic series: repetition of **AE1**
(4) **RES2** – 2nd resonant series: repetition of **RES1**

For the analysis of frequency jumps, narrow-band spectrograms (500 ms segments using a Gaussian window, 50% overlap) were obtained from the microphone signal after pre-emphasizing (+6 dB per octave). The spectrogram viewing range was set to 0–450 Hz to visualize f_o and the lowest harmonics.

We also extracted the f_o contour from the microphone signals, using the SWIPE' algorithm designed by Camacho et al. [14]. The signal was divided into 100 ms segments with a 50% overlap between segments, and one value of f_o was estimated for every segment. From the f_o contour, we manually selected the values directly before and after each frequency jump to estimate its extent.

3. Results

Two of the five larynges - larynx #1 (aged 55 years) and #2 (aged 69 years) - exhibited consistent frequency jumps during the elongation-shortening sweeps for both anechoic and resonant conditions. The remaining three larynges exhibited none or sparse frequency jumps that were not repeatable and did not provide reliable measurable data. Therefore, our pilot results are limited to two of the five larynges. The values of the air flow were at approximately 320 mL/s for larynx #1 and 270 mL/s for larynx #2. The mean subglottal pressure ranged from 6.6 to 26.7 cm H_2O for larynx #1, and from 6.6 to 14.9 cm H_2O for larynx #2. These values fall within the usual ranges of flows and pressures observed in excised human larynges [15].

Spectrograms of the sweeps for larynx #1 and #2 are shown on Fig. 3, showing consistent f_o jumps (indicated by blue arrows in Fig. 3) during the upward and downward frequency sweeps for both the anechoic and the resonant subglottal condition. The number of jumps and their appearance were similar for the anechoic and resonant conditions; no additional jumps emerged when the resonant subglottal tract was present. In both larynges and both conditions (resonant vs. anechoic), systematic hysteresis effects were observed: the upward jumps always occurred at higher frequencies than the downward jumps, which is consistent with the observations of other studies [2,7,16]. In addition, regardless of the subglottal conditions, secondary jumps appeared occasionally for both larynges, which were not reproducible. Therefore, we did not consider that latter class of phenomena in our analyses.

Visually, the spectrograms did not reveal any clear difference in the frequency jumps between anechoic and resonant conditions. To find out more, quantitative analysis was performed. The values of f_o measured before and after each jump are shown in Fig. 4, for larynx #1 and #2. Within each condition, the repeated experiments showed considerable variability as regards the initial and terminating frequency of the jumps.

Considering the upward jumps, in larynx #1 (Fig. 4a), the starting frequency was between 120 and 170 Hz in the anechoic and between 125 and 165 Hz in the resonant condition. The terminating frequency was between 215 and 250 Hz in the anechoic and between 240 and 270 Hz in the resonant condition. In larynx #2 (Fig. 4c), the starting jump frequency was slightly lower, i.e., between 110 and 125 Hz in the anechoic and between 110 and 130 Hz in the resonant condition. The terminating frequency was found in a range of 240–280 Hz in the anechoic condition (higher than in larynx #1, with 10 Hz overlap) and 205–235 Hz (lower than in larynx #1) in the resonant condition. Altogether, the upward jump frequencies were different between the two larynges, and the subglottal resonances noticeably changed the terminating frequencies of the jumps.

Looking at the downward jumps, in larynx #1 (Fig. 4b), the starting frequency was between 160 and 185 Hz in the anechoic and between 160 and 210 Hz in the resonant condition. The terminating frequency was between 110 and 130 Hz in both the anechoic and resonant conditions. In larynx #2 (Fig. 4d), the starting frequency of the downward jumps was between 170 and 245 Hz in the anechoic and between 155 and 260 Hz in the resonant condition (larger variability than in larynx #1). The terminating frequency was between 115 and 130 Hz in both anechoic and resonant conditions, similarly to larynx #1. Overall, the differences between the two larynges and between the two conditions were smaller in the downward jumps than in the upward jumps.

The two series of tasks in identical conditions (i.e., AE1 vs. AE2, and RES1 vs. RES2) were used to check the repeatability of the events over time. Mostly, no systematic differences were found in the frequency jumps between the two series (see the full versus dotted lines in Fig. 4). Nevertheless, in two out of the eight cases, the values of f_o both before and after the frequency jumps were slightly lower in the second than in the first series: (a) during the upward frequency jumps for larynx #1 in the anechoic condition (Fig. 4a); and (b) during the downward frequency jumps for larynx #2 in resonant condition (Fig. 4d). For larynx #1, phonation was lost after reaching high values of f_o during some of the sweeps, and no downward jump was observed during the third sweep of the RES2 series (see Fig. 3d).

4. Discussion and conclusion

This study presents the very first experiments on f_o jumps in excised human larynges conducted in acoustically anechoic conditions. We prove that the jumps occur in excised larynges in anechoic conditions. Furthermore, we have not observed the resonant subglottal tract to stimulate more numerous frequency jumps compared to anechoic conditions, confirming that those jumps are primarily caused by inherent laryngeal properties excited by the airflow rather than by acoustic interactions with subglottal and/or supraglottal tracts. These inherent properties differ among larynges: we found differences between our two larynges especially during upward jumps (see, e.g., Fig. 4a,c).

Besides the differences between the larynges, there was considerable variability of the frequency jumps within each larynx. This

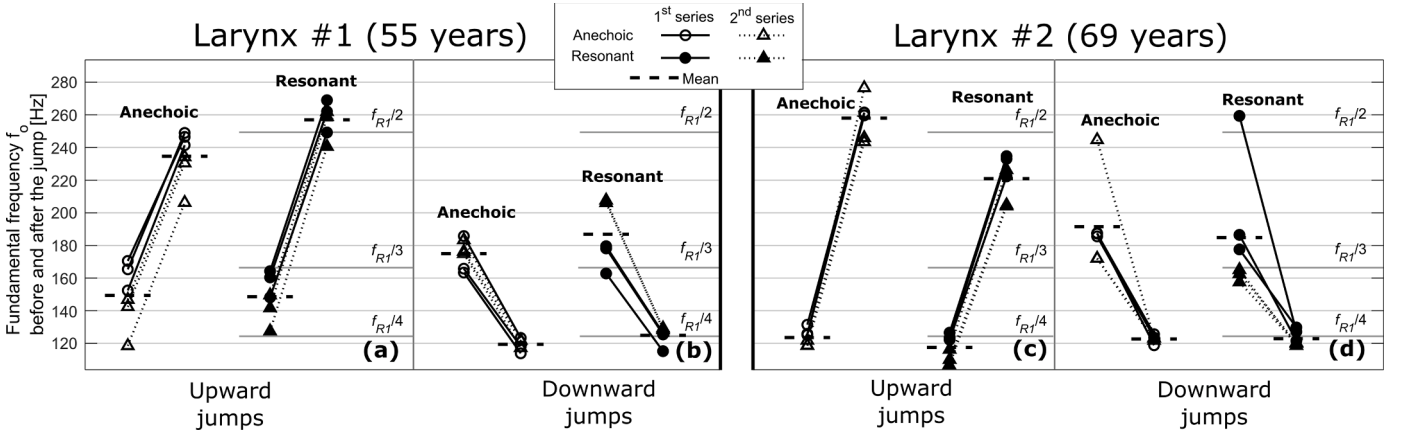


Fig. 4. Fundamental frequency (f_o) before and after each frequency jump for larynx #1 and #2. The circles correspond to the first, and the triangles to the second series of frequency sweeps. For the resonant cases, we also indicate the integer divisions of the first subglottal resonance frequency ($f_{R1} \approx 500$ Hz), $f_{R1}/2$, $f_{R1}/3$, and $f_{R1}/4$. (a) larynx #1, upward jumps; (b) larynx #1, downward jumps; (c) larynx #2, upward jumps; (d) larynx #2, downward jumps.

jump variability was not only present in anechoic, but also in resonant conditions. Because the subglottal resonances were held constant, this variability suggests that the f_o values during the jumps were not strongly entrained to the subglottal resonance frequencies. The potential entrainment frequencies are indicated in Fig. 4 as horizontal lines marked as $f_{R1}/2$, $f_{R1}/3$ and $f_{R1}/4$, corresponding to the attraction of the 2nd, 3rd and 4th harmonics ($2f_o$, $3f_o$ and $4f_o$) to the lowest subglottal resonance frequency f_{R1}. The initial and terminating f_o frequently did not coincide with these lines. This contrasts with the findings of Zhang et al. [5], who used a physical model of the VFs coupled to an extendable ST and found the f_o to be attracted to the first four subglottal resonance frequencies f_{R1}, f_{R2}, f_{R3}, and f_{R4}. However, our ST was different in that it approximated the values of the first subglottal resonance frequency f_{R1} found *in vivo* (approximately 500 Hz), which was considerably higher than the values used by Zhang et al. [5]. Consequently, our f_os were never in the vicinity of f_{R1}, which can explain the different results from [5].

When comparing the jump characteristics between anechoic (AE) and resonant (RES) conditions, we observed that the jump variability within a single (AE or RES) condition was larger or comparable to the differences between AE and RES conditions in 7 out of the 8 cases shown in Fig. 4. This again suggests that the subglottal resonances did not play a primary role for the jump characteristics. Nevertheless, during the upward jumps the terminating f_o values were noticeably different between AE and RES conditions (Fig. 4a,c). In one case, i.e., the upward jumps in larynx #2 (Fig. 4c), there was about a 37 Hz difference between the AE and RES mean value, which was larger than the inter-attempt variability (33 Hz in AE condition and 31 Hz in RES condition), and there was no overlap between the AE and RES values. Spencer et al. [17] reported similar findings *in vivo* when the subglottal resonance frequencies were changed through inhalation of helox gas (a mixture of 20% oxygen and 80% helium): one out of four subjects exhibited an average increase of the starting and terminating frequencies of the jumps in these conditions. This indicates that the acoustic subglottal resonances also have an influence on the jump characteristics. However, such an influence appears to be secondary and not as dominant as that of the inherent properties of the larynges.

Generally, the acoustic damping properties of the ST influence the strength of the interaction between the subglottal acoustics and the VF oscillations. In our experiment, the resonance bandwidth of our ST (for $f_{R1} = 500$ Hz) was about 43 Hz (Q factor = 0.086) (see Fig. 2), which is considerably smaller than the subglottal resonance bandwidths of > 100 Hz (Q factor > 0.2) measured *in vivo* [13]. A smaller resonance bandwidth (smaller Q factor) reflects less subglottal damping. Therefore, our subglottal resonance effects are expected to be stronger than those occurring *in vivo*. Despite this, our observed subglottal resonance effects were still not as prominent as the vocal tract resonance effects on the frequency jumps observed *in vivo* [3,4,16,18]. This corroborates the findings from simulation experiments of Tokuda et al. [16], who reported that the influence of the subglottal resonances on potential frequency jumps is rather small in comparison with the influence of the vocal tract acoustic resonances.

The data presented in this pilot study stem from two excised human larynges. The other three investigated larynges exhibited either none or very sparse f_o jumps that were not reproducible. We attribute the lack of f_o jump occurrences to the high age of the subjects from which the larynges were collected (66–77 years). Such larynges have the tendency to stay in falsetto-like vibratory regime and do not switch to the chest-like regime. Due to ethical constraints, we did not collect more human larynges for these pilot investigations, which is a limitation of this study. Nevertheless, our results provide a definite proof that the f_o jumps can occur in excised human larynges within acoustically anechoic conditions and therefore are primarily source-induced. It concurs with the explanation that the jumps result from different laryngeal mechanisms [19] and from VF vibratory mode changes [8,20]. These novel experimental findings can be used for tuning the properties of mathematical models aiming at replicating the nonlinear-dynamic behavior of the human vocal apparatus.

CRediT authorship contribution statement

Hugo Lehoux: Methodology, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization. **Christian T. Herbst:** Conceptualization, Investigation, Writing – review & editing. **Martin Dobiáš:** Resources, Writing – review & editing. **Jan G. Švec:** Conceptualization, Methodology, Investigation, Writing – review & editing, Supervision, Funding acquisition.

Declaration of Competing Interest

None.

Data Availability

Data will be made available on request.

Acknowledgments

This work was supported by the Czech Science Foundation (GA CR) project no. 19-04477S (to JGS) and by the Land Salzburg research grant no. 20204-WISS/245/2-2019 (to CTH). The stipend for doctoral studies of HL was provided by the Palacký University internal grant IGA_PrF_2022_029.

References

[1] H. Herzel, D. Berry, I.R. Titze, M. Saleh, Analysis of vocal disorders with methods from nonlinear dynamics, J. Speech Lang. Hear. Res. 37 (5) (1994) 1008–1019, https://doi.org/10.1044/jshr.3705.1008.

[2] J.G. Švec, H.K. Schutte, D.G. Miller, On pitch jumps between chest and falsetto registers in voice: data from living and excised human larynges, J. Acoust. Soc. Am. 106 (3) (1999) 1523–1531, https://doi.org/10.1121/1.427149.

[3] I.R. Titze, T. Riede, P. Popolo, Nonlinear source–filter coupling in phonation: vocal exercises, J. Acoust. Soc. Am. 123 (4) (2008) 1902–1915, https://doi.org/10.1121/1.2832339.

[4] L. Wade, N. Hanna, J. Smith, J. Wolfe, The role of vocal tract and subglottal resonances in producing vocal instabilities, J. Acoust. Soc. Am. 141 (3) (2017) 1546–1559, https://doi.org/10.1121/1.4976954.

[5] Z. Zhang, J. Neubauer, D.A. Berry, The influence of subglottal acoustics on laboratory models of phonation, J. Acoust. Soc. Am. 120 (3) (2006) 1558–1569, https://doi.org/10.1121/1.2225682.

[6] D.A. Berry, H. Herzel, I.R. Titze, B.H. Story, Bifurcations in excised larynx experiments, J. Voice 10 (2) (1996) 129–138, https://doi.org/10.1016/S0892-1997(96)80039-7.

[7] I.T. Tokuda, J. Horáček, J.G. Švec, H. Herzel, Comparison of biomechanical modeling of register transitions and voice instabilities with excised larynx experiments, J. Acoust. Soc. Am. 122 (1) (2007) 519–531, https://doi.org/10.1121/1.2741210.

[8] B. Geng, M. Movahhedi, Q. Xue, X. Zheng, Vocal fold vibration mode changes due to cricothyroid and thyroarytenoid muscle interaction in a three-dimensional model of the canine larynx, J. Acoust. Soc. Am. 150 (2) (2021) 1176–1187, https://doi.org/10.1121/10.0005883.

[9] H. Lehoux, V. Hampala, J.G. Švec, Subglottal pressure oscillations in anechoic and resonant conditions and their influence on excised larynx phonations, Sci. Rep. 11 (1) (2021) 28, https://doi.org/10.1038/s41598-020-79265-3.

[10] Act no. 372/2011 on Health Services and Conditions of Their Provision. [Zákon 372/2011 Sb o zdravotních službách a podmínkách jejich poskytování.], in: Collection of Law of the Czech Republic, vol. 131, 2011, pp. 4777–4782.

[11] M. Garcia, C.T. Herbst, Excised larynx experimentation: history, current developments, and prospects for bioacoustic research, Anthropol. Sci. 126 (1) (2018) 9–17, https://doi.org/10.1537/ase.171216.

[12] I.R. Titze, Experiments With Excised larynges. in: The Myoelastic Aerodynamic Theory of Phonation, National Center for Voice and Speech, Denver CO and Iowa City IA, 2006, pp. 1–62.

[13] G. Fant, K. Ishizaka, J. Lindqvist, J. Sundberg, Subglottal formants, STL-QPSR 1 (1972) 1–12, 1972.

[14] A. Camacho, J.G. Harris, A sawtooth waveform inspired pitch estimator for speech and music, J. Acoust. Soc. Am. 124 (3) (2008) 1638–1652, https://doi.org/10.1121/1.2951592.

[15] J. Van den Berg, T.S. Tan, Results of experiments with human larynxes, Pract. Oto-Rhino-Laryng. 21 (6) (1959) 425–450, https://doi.org/10.1159/000274240.

[16] I.T. Tokuda, M. Zemke, M. Kob, H. Herzel, Biomechanical modeling of register transitions and the role of vocal tract resonators, J. Acoust. Soc. Am. 127 (3) (2010) 1528–1536, https://doi.org/10.1121/1.3299201.

[17] M.L. Spencer, I.R. Titze, An investigation of a modal-falsetto register transition hypothesis using helox gas, J. Voice 15 (1) (2001) 15–24, https://doi.org/10.1016/S0892-1997(01)00003-0.

[18] Y. Uezu, T. Kaburagi, A measurement study on voice instabilities during modal-falsetto register transition, Acoust. Sci. Technol. 37 (6) (2016) 267–276, https://doi.org/10.1250/ast.37.267.

[19] B. Roubeau, N. Henrich, M. Castellengo, Laryngeal vibratory mechanisms: the notion of vocal register revisited, J. Voice 23 (4) (2009) 425–438, https://doi.org/10.1016/j.jvoice.2007.10.014.

[20] P. Schlegel, D.A. Berry, D.K. Chhetri, Analysis of vibratory mode changes in symmetric and asymmetric activation of the canine larynx, PLoS One 17 (4) (2022), e0266910, https://doi.org/10.1371/journal.pone.0266910.

Supplement C: Paper III

Laryngeal and Acoustic Analysis of Chest and Head Registers Extended Across a Three-Octave Range: A Case Study.

H. LEHOUX, L. POPEIL, J. G. ŠVEC

Journal of Voice (2022, in press)

doi: 10.1016/j.jvoice.2022.02.014

Laryngeal and Acoustic Analysis of Chest and Head Registers Extended Across a Three-Octave Range: A Case Study

*Hugo Lehoux, †Lisa Popeil, and *Jan G. Švec, *Olomouc, Czech Republic, and †Los Angeles, California

Summary: Voice registers are assumed to be related to different laryngeal adjustments, but objective evidence has been insufficient. While chest register is usually associated with the lower pitch range, and head register with the higher pitch range, here we investigated a professional singer who claimed an ability to produce both these registers at every pitch, throughout her entire singing range. The singer performed separated phonations alternating between the two registers (further called chest-like and head-like) at all pitches from C3 (131 Hz) to C6 (1047 Hz). We monitored the vocal fold vibrations using high-speed video endoscopy and electroglottography. The microphone sound was recorded and used for blind listening tests performed by the three authors (insiders) and by six "naive" participants (outsiders). The outsiders correctly identified the registers in 64% of the cases, and the insiders in 89% of the cases. Objective analysis revealed larger closed quotient and vertical phase differences for the chest-like register within the lower range below G4 (<392 Hz), and also a larger closed quotient at the membranous glottis within the higher range above Bb4 (>466 Hz), but not between Ab4-A4 (415-440 Hz). The normalized amplitude quotient was consistently lower in the chest-like register throughout the entire range. The results indicate that that the singer employed subtle laryngeal control mechanisms for the chest-like and head-like phonations on top of the traditionally recognized low-pitched chest and high-pitched head register phenomena. Across all pitches, the chest-like register was produced with more rapid glottal closure that was usually, but not necessarily, accompanied also by stronger adduction of membranous glottis. These register changes were not always easily perceivable by listeners, however.

Key Words: Voice registers−Vocal fold oscillation−Singing range−High-speed videoendoscopy−Electroglottography.

INTRODUCTION

The division of the human voice into different regions is a widely accepted but highly controversial concept in the voice research and singing community. Such regions are called "registers", a term that originated as an analogy to church organ registers.[1] The vocal registers can have many different names, definitions, and frequency ranges according to the language or the teaching methods: in 1963, Mörner et al, reported more than a hundred different terms to describe registers.[2] Nevertheless, it is most commonly accepted that there are two main registers used in singing, and they are named "chest" (or "modal") and "head" (or "falsetto") registers. There is currently no strictly objective method to differentiate those registers, but singers or singing teachers appear to be able to differentiate them from their singing or teaching experience by recognizing acoustic characteristics or proprioceptive sensations. A sudden register change can also be recognized directly from a transition event, such as a voice break or a pitch jump. Such events typically occur at the transition between the chest and head registers, for example when singing a glissando.[3]

Registers are usually named after their perceptual or proprioceptive characteristics. For example, the terms "chest" and "head" are associated with a sensation of sympathetic vibrations in the chest and head areas respectively. Despite those characteristics, these two registers are assumed to be associated mostly with different laryngeal adjustments, which produce different vocal fold vibratory patterns.[4] Such laryngeal adjustments are sometimes referred to as "laryngeal mechanisms" and numbered from low pitch to high pitch (M0, M1, M2 and M3),[5] where M1 and M2 are usually associated with the chest and the head registers, respectively. In order to observe the activity of the laryngeal muscles when using different registers, Hirano et al,[6] performed electromyographic (EMG) measurements. They found that the thyroarytenoid (TA) muscle (also called the vocalis muscle) was generally more active in the chest register than in the head register. Similar findings were reported also by the studies of Baer et al,[7] and Kochis-Jennings et al,[8] although the TA dominance for the chest register appears questionable at the fundamental frequencies above 300 Hz.[9]

The electroglottographic (EGG) signal[5,10,11] and its relation to registers has also been analyzed: it has been found that its amplitude generally decreases and its open quotient increases when singers shift from the chest to the head register. Laryngeal imaging has finally proven to be a useful tool to observe the vibrations of the vocal folds, through simple laryngoscopy,[12,13] videokymography[14] or high-speed videoendoscopy.[15-22] General findings mostly agree

Accepted for publication February 11, 2022.
Conflict of interest: The authors declare that they have no financial conflicts of interest to disclose.
The work was supported by the Czech Science Foundation (GA CR) project no. 19-04477S. The stipend for doctoral studies of HL was provided by the Palacký University internal grant IGA_PrF_2021_017.
From the *Voice Research Lab, Department of Experimental Physics, Faculty of Science, Palacký University, Olomouc, Czech Republic; and the †Voiceworks, Los Angeles, California.
Address correspondence and reprint requests to Jan G. Švec Voice Research Lab, Department of Experimental Physics, Faculty of Science, Palacký University, 17. listopadu 12, 771 46, Olomouc, Czech Republic. E-mail: jan.svec@upol.cz
Journal of Voice, Vol. ■■, No. ■■, pp. ■■−■■
0892-1997

https://doi.org/10.1016/j.jvoice.2022.02.014

on a firmer, longer vocal fold contact in the chest than in the head register.

Registers are often associated with pitch (a perceptual correlate to fundamental frequency f_o, in singing often expressed by musical notes): the chest register is mostly used for comfortable speech and lower notes and the head register is used for higher notes. There exists an overlap region where both these registers can be used, however. Therefore, previous works have studied registers either by comparing parameters at the same pitch in this overlap region, or by analyzing the transition occurring in ascending or descending glissandi.

While the register transitions are traditionally smoothed out in classical western operatic singing, some non-operatic singing styles, such as Country and Western yodeling, or Pop "falsetto flips", explore the register transitions for artistic purposes. In 2008, the Australian singer Mal Webb published a short video describing a special singing technique which he calls "sideways yodeling".[23] He defined this technique as a sudden change of register (from chest to head or from head to chest) while keeping the same pitch, or even switching from a higher chest-register pitch to a lower head-register pitch, which contrasts with the traditional yodeling techniques. Interestingly, this singer claimed that the chest and head registers are in fact not limited by the pitch, as he demonstrates using the chest register up to high notes like E5 = 659 Hz and using the head register down to low notes like Ab2 = 104 Hz.

In this study, we investigated a female professional singer and singing pedagogue (co-author LP) who took inspiration from Mal Webb's "sideways" register transitions and extended it to her entire singing range. In other words, she claimed that she can alternatively use both chest and head registers regardless of pitch, according to her own kinesthetic and auditory perceptions, and therefore that her chest-to-head-register-overlap region was extended throughout her full pitch range. We decided to perform a detailed experimental laryngeal investigation of this subject using laryngeal high-speed videoendoscopic, electroglottographic (EGG) and audio recordings of the produced vocal sound. Besides quantitative analyses of the recorded signals, we also performed listening tests to assess the perceptual distinguishability of the two vocal registers across the whole pitch range.

During the course of the study, we ran into terminological problems with assigning proper names to the two registers investigated in this study. The singer initially used the terms "M1" and "M2", but we decided to avoid these since we had no proof that the two registers are two distinct vibratory mechanisms separated by a clear transition event, as understood by the scientists who introduced the M1, M2 register terminology (N. Henrich, personal communication). The traditional terms "chest" and "head" registers, on the other hand, we considered potentially confusing, since they are commonly associated with low and high pitches. We finally opted for the terms "chest-like" and "head-like" which are used in the rest of this article to refer to the singer's intended registers. The terms "chest" and "head" are used here to refer to the registers described previously in literature.

Based on the previous findings reporting more TA muscle activity[6-9] and acoustically richer harmonic structure of the sound in chest than in the head register,[24,25] we formulated the following hypotheses for our study:

(1) *The vocal folds will have a longer and more pronounced contact in the chest-like register than in the head-like register.* This hypothesis is based on the knowledge that the TA muscle bulges the membranous portion of the vocal folds[26,27] allowing firmer membranous adduction.[28,29]
(2) *The vertical phase differences between the vibrations of the lower and upper margins of the vocal folds will be larger in the chest-like register than in the head-like register.* This hypothesis stems from the knowledge that the TA muscle makes the vocal fold thicker vertically,[26] thus allowing larger differences between the motions of the upper and lower margins.
(3) *The amplitude of vibration of the lower margin of the vocal folds will be larger and, consequently, the maximum glottal closing speed will be higher in the chest-like register than in the head-like register.* This hypothesis takes into account that the more adducted lower margin of the vocal folds likely vibrates with larger amplitude, which presents itself through larger maximum area declination rate.[30]
(4) *There will be more high-frequency energy in the sound spectrum of the chest-like compared to the head-like register.* This hypothesis is based on previous observations reporting a flatter spectral slope of the sound source in chest than in head register.[24,25]

METHODS

Experimental setup

A female professional singer and singing pedagogue (co-author LP) with more than 40 years of experience with performing, recording, and teaching singing served as a subject for this study. She performed short (between 0.9 and 2.2 s) sustained phonations on all musical notes (semitones) throughout the range of three octaves from C3 (131 Hz) to C6 (1047 Hz), first in the chest-like then (after a short pause) in the head-like register, using the "laryngoscopic vowel" similar to the neutral vowel /ə:/. No recording was performed on one pitch (Ab3 = 208 Hz), however, as it was omitted by mistake. We also asked the singer to perform continuous phonations, alternating between the chest-like and head-like registers without a pause, to verify findings related to the changes in laryngeal configuration and to check for discontinuities during the register transitions. This second task was only performed at pitches Db3 (139 Hz), Bb3 (233 Hz), F4 (349 Hz), C5 (523 Hz), F5 (698 Hz) and Bb5 (932 Hz) to reduce the excessive duration of the experiment (c. 10-minutes waiting time was needed for data storage before starting a new recording, making the experiment time-demanding). We monitored the vocal fold vibrations

with a rigid 90° endoscope (8707 DA, 10 mm, Karl STORZ, Germany) connected to a high-speed camera (FASTCAM Mini AX100 54, 16GB memory, Photron, Japan). We used a 350 W xenon light source (Hawk HW-300A) connected to the endoscope via a light-conducting cable (P2551A HAWK, 5 × 2500 mm). All the phonations were registered by the camera at 7200 frames per second (fps), with the resolution of 384 × 576 pixels except the high-pitched phonations from D5 (587 Hz) to C6 (1047 Hz) which were recorded at the increased high-speed rate of 13600 fps. The videos were saved as uncompressed AVI files.

Simultaneously, we registered the radiated sound through a small condenser microphone (Isomax 2 omnidirectional microphone, Countryman Associates, USA) placed on top of the camera as depicted in Figure 1. During the phonation, the mouth-to-microphone distance was measured to be 22 cm. Furthermore, we simultaneously registered the EGG signal with electrodes placed at the vocal fold level using the GLOTTAL ENTERPRISES EG2-PC EGG device. Finally, we used an impulse signal generated by the camera to synchronize the audio and the video events. The microphone, EGG and synchronization signals were sampled at 48 kHz, 24 bits resolution using the Fireface 400 soundcard (RME, Germany) connected to a second computer, which recorded the signals using the Audacity software[31] and saved them as synchronized uncompressed mono WAV files. The high-speed video recordings were limited by the camera storage memory to about 7 seconds at 7200 fps and about 4 seconds at 13600 fps. Each recording contained at least one phonation of each register, and the intended register of every phonation according to the subject's own proprioception was noted.

Listening test

Listening tests were performed to determine whether the two different registers could be audibly distinguished. For this purpose, the chest-like and head-like phonations from the microphone signal, including the phonation onsets and offsets, were manually separated and saved as individual WAV files. There were 72 samples in total, containing two audio samples per each pitch from C3 = 131 Hz to C6 = 1047 Hz (except for the pitch Ab3 = 208 Hz as previously mentioned in section 2.1), i.e., 36 pitches in total: one in the chest-like register and one in the head-like register. Each sample was individually normalized by setting the maximum absolute value of the signal to 1. A total of 9 evaluators independently performed the listening tests. To assess the repeatability of the perceptual evaluations, each listener completed the test twice with at least 1-week interval between the two attempts to minimize the memory effect. The evaluators were divided in two groups − "insiders" and "outsiders". The "insiders" group contained the three authors of this study who were familiar with the study design and knew that each sample was intended to be either in the chest-like or in the head-like register, according to the professional singer. The six listeners from the "outsiders" group were recruited among acquaintances of the three authors. Two of them were singing teachers and the others were singers in the university student choir. They were all familiar with the sound of chest and head registers but were not familiar with the study design. They were given detailed instructions on how to perform the listening test (the instructions are available in the appendix). The listeners self-rated their ability to produce and to recognize both registers, on a scale from 0 (not experienced) to 10 (expert) (see the appendix). Table 1 lists the singing experience in years and the claimed ability to recognize both registers, for each participant.

The listening tests were performed using the Visor software[32,33] which allows visual sorting and rating of audio samples. The software interface is composed of icons (one icon for each audio sample) the listeners can right-click on, to listen to the sample as many times as needed (see Figure 2). The listeners were asked to rate and order the audio samples on a continuous scale going from "chest-like" phonation to "head-like" phonation, by moving the samples up (more "head-like") or down (more "chest-like"). The listener could also freely arrange the samples horizontally according to the pitch, but the horizontal position was not saved by the software and therefore had no impact on the test result. In addition, the placement of each sample was fully randomized every time the test started. When the

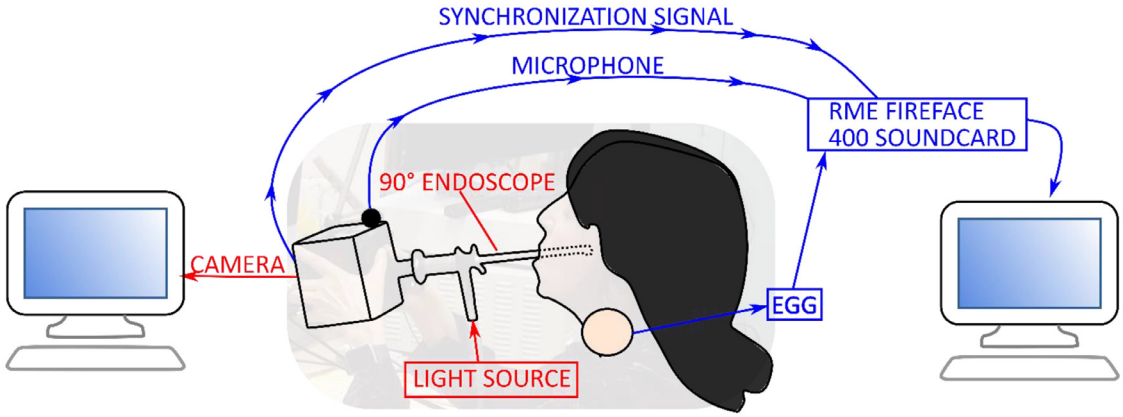


FIGURE 1. Experimental setup for the experiments. (For color version of this figure, the reader is referred to the Web version of this article.)

TABLE 1.
General Information About the Participants in the Listening Tests

	Singing experience (years)	Claimed ability to recognize chest and head registers E_R (0: not experienced, 10: expert)
Outsiders subjects		
O1	12	7
O2	20+	9
O3	14	2
O4	18	8.5
O5	12	4
O6	11	4
Insiders subjects		
HL	4	7
JGS	20+	8
LP	20+	10

test was finished, the software saved the results as a text file which we subsequently analyzed using custom MATLAB scripts.

The rating scale was continuous between "chest-like" and "head-like"; the score 0 meaning the "pure chest" register and 1000 meant the "pure head". For the purpose of the analysis, we defined three possible outcomes for the results per pitch in the following way:

- Correct: the sample intended in the chest-like register was rated more than 30 points lower (more chest-like) than the sample intended in the head register, for the same pitch. This outcome was represented by the "score" of +1.
- Similar: the sample intended in the chest-like register was rated at most 30 points lower or higher than the sample intended in the head-like register, for the same pitch. This outcome was represented by the "score" of 0.
- Incorrect: the sample intended in the chest-like register was rated more than 30 points higher (more head-like) than the sample intended in the head-like register, for the same pitch. This outcome was represented by the "score" of -1.

For each group and for each pitch, the sum of the "scores" was defined as the "agreement score", which was expressed per pitch and per group, and could have a value ranging from $-n$ (100% disagreement with the intended register) to $+n$ (100% agreement with the intended register), where n is the number of tests in the corresponding group (12 for the outsiders group and 6 for the insiders group).

Processing of the high-speed videos

For each video file, we obtained two 100-ms-long digital kymograms (one for the chest-like phonation and one for the head-like phonation) from the position at the middle of the membranous glottis, perpendicularly to the glottal axis, using the software ImageJ and Python scripts developed by Christian T. Herbst.[34] These kymograms were segmented by the VKGAnalyzer software[35] to obtain the oscillation contours of the left and right vocal folds. These contours were used for analysis using custom Matlab scripts. We also segmented the

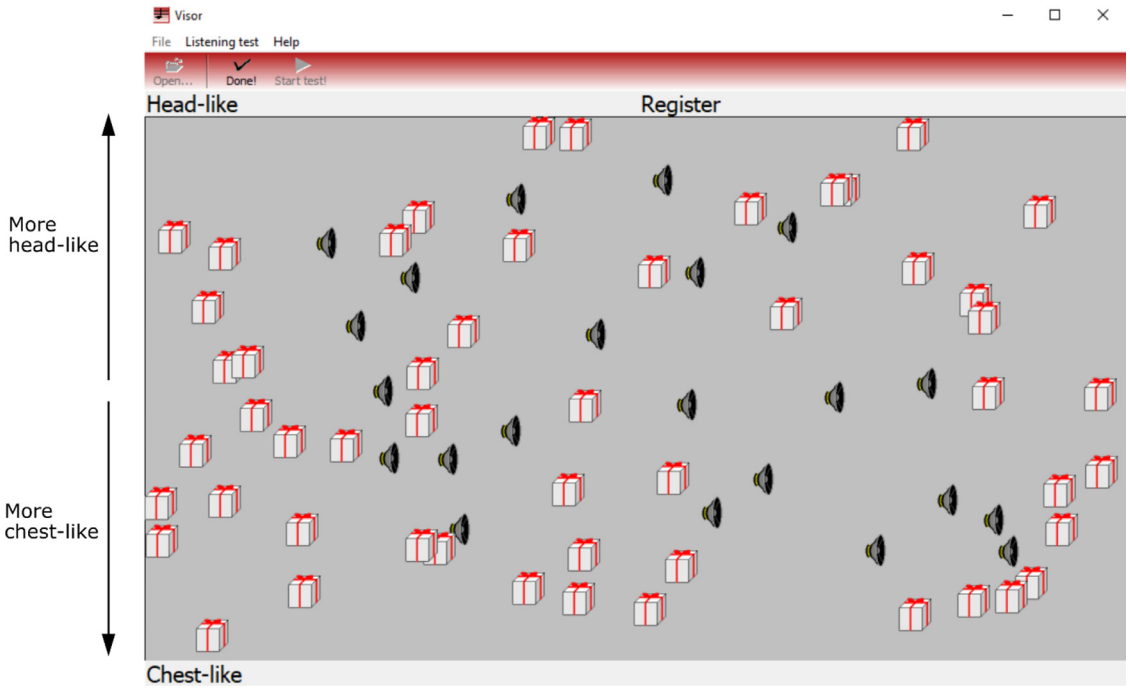


FIGURE 2. Screenshot of the Visor window, where all the sample icons appear. The loudspeaker icons represent samples which have been played at least once, and the 'gift' icons show the samples which have never been played.

full-sized high-speed videos using Glottal Analysis Tools 2018 (Erlangen, Germany) to extract the glottal area waveform (GAW)[36-38] and to obtain Phonovibrograms (PVGs) of the vocal fold edges.[39] PVGs are a measure of the distance of the glottal edge from the glottal midline for each vocal fold, segmented into 256 positions along the longitudinal axis of the glottal midline. Glottal Analysis Tool was also used to automatically detect the vocal fold oscillation cycles. For the tones between Ab4 (466 Hz) and C6 (1047 Hz), we performed the segmentation on the sections of the high-speed videos corresponding to the 100 ms windows of the obtained kymograms. The segmentation was done similarly for the tones below Ab4, but for those we also included 100 ms just before and after the part corresponding to the obtained kymograms, for a total duration of 300 ms per segmentation, in order to obtain more cycles of vibration.

Processing of the microphone and EGG signals

For the analysis of the microphone and EGG signals, the synchronized microphone and EGG channels were edited to separate the chest-like and head-like phonations. The phonation onsets and offsets were removed here in order to keep only the stable phonation sections. This resulted in a total of 144 samples (72 samples for each signal): one per pitch and per register. The microphone signal was calibrated in pascal using the 2A method from Švec and Granqvist[40] and the associated MATLAB scripts.[41]

In order to visually compare the spectra of all phonations, the power spectrum of each microphone sample was calculated from 100 Hz to 5000 Hz at 24 bands per octave using the "poctave" Matlab built-in function, after applying a pre-emphasis filter (+6 dB per octave). For each phonation, the spectrum was normalized by dividing the power of all the spectral components by the power of the maximum spectral component. Then gray-scale coding was applied so that the individual spectra at all pitches could be presented visually as spectrogram-like images (refer to Figure 11B,C in section 3.3). For calculation of the spectral parameters, the pre-emphasis filtering and the spectral normalization were not applied.

Calculation of the parameters

According to the hypotheses formulated in the introduction, we calculated several parameters for which we expected to see a substantial difference between the chest-like and head-like registers. In relation to hypothesis (1), the selected parameters were: the closed quotient from the PVG signal (CQ_{PVG}) and the EGG amplitude. The parameter related to hypothesis (2) was the vertical phase difference, ie, the phase delay between the movements of the lower and upper vocal fold margins. To test hypothesis (3), we calculated the normalized amplitude quotient from the GAW (NAQ_{GA}), the speed quotient from digital kymograms (SQ_{kym}), and the closing quotient from digital kymograms (CgQ_{kym}). To account for hypothesis (4), we calculated the spectral balance from the microphone signal (SB) and the difference between the levels of the first and second harmonics (L_{H1}-L_{H2}) from the microphone signal. The calculation of all those parameters is detailed below.

Parameters obtained from laryngeal high-speed videos

Closed quotient (CQ_{PVG}) – this parameter provides information on the relative glottal closure duration during phonation. The CQ_{PVG} is expected to be longer in chest than in head register. From the PVG, we obtained the closed quotient for each cycle by dividing the number of frames where the glottis was closed (i.e., the distance between both vocal folds was 0) by the duration of the cycle (in frames). The CQ_{PVG} was calculated in this way at each of the 256 positions along the longitudinal axis of the glottal midline and was then averaged over the cycles, to obtain one mean value and its standard error per longitudinal position (referred to as the cycle-averaged local CQ_{PVG}). To account for the possible anterior-posterior differences in glottal adduction, in every recording we divided the glottis along the glottal axis into three sections: anterior, middle and posterior, comprising respectively 85, 86, and 85 positions, and selected the median value of the cycle-averaged local CQ_{PVG} in these three sections, similarly as done by Sielska-Badurek et al, for the open quotient.[42]

Normalized amplitude quotient (NAQ_{GA}) – this parameter represents an idealized relative duration of the closing phase assuming the glottis moves with the maximum observed velocity[43,44] for the whole duration of the closing phase. It was calculated from the maximum area declination rate[30,36,37] using the following equation:

$$NAQ_{GA} = \frac{max(GAW) - min(GAW)}{MaximumAreaDeclinationRate * T} \quad (1)$$

where T is the cycle duration. We expected NAQ_{GA} to be lower in the chest-like than in head-like register, reflecting a more rapid glottal closing movement in the chest-like register.

Vertical phase difference (VPD_{kym}) – this parameter is an estimation of the phase delay of the upper margin of the vocal folds behind the motion of the lower vocal fold margin and is related to the mucosal wave velocity on the medial vocal fold surface.[45,46] The VPD_{kym} was expected to be larger in chest-like than in head-like register. We extracted the vocal fold contours from the digital kymograms segmented by the VKGanalyzer software.[35] We estimated the VPD_{kym} between the lower and upper vocal fold margins using the sinusoid fitting method detailed in Jiang et al,[47] and we obtained one value per pitch and per vocal fold.

Speed quotient (SQ_{kym}) – this parameter compares the average speed of the opening and closing phases of the glottal cycle. It is defined as the ratio between the opening phase duration t_{og} and the closing phase duration t_{cg}:[48]

$$SQ_{kym} = \frac{t_{og}}{t_{cg}} \quad (2)$$

The SQ_{kym} is assumed to reflect the ratio between the amplitude of the lower versus the upper margin, increasing values indicating larger amplitude of the lower versus the upper margin.[30] For chest-like register, we expected the lower margin amplitude to be larger and therefore also the SQ_{kym} to be larger than for the head-like register. We obtained the SQ_{kym} using the vocal fold contour extracted from digital kymograms.

Closing quotient (CgQ_{kym}) – this parameter measures the relative duration of the closing phase with regards to the cycle duration; it was defined as the closing phase duration t_{cg} divided by the cycle duration T:[49,50]

$$CgQ_{kym} = \frac{t_{cg}}{T} \quad (3)$$

The CgQ_{kym} is expected to decrease when the closing movement becomes more rapid. We expected it to be smaller for chest-like than for head-like register. Similar to the speed quotient and the vertical phase difference, the closing quotient was calculated using the vocal fold contours, extracted from the digital kymograms via VKGanalyzer.

Acoustic parameters

Sound pressure level (SPL) - we calculated the SPL of the calibrated signal using the "C" frequency weighting and the "special" time weighting (50 ms moving average), and we kept one value every 40 ms (1920 samples). As previously mentioned, the mouth-to-microphone distance was approximately 22 cm.

Spectral balance (SB) – this parameter served as a qualitative indication of the voice sound quality. The spectral balance was calculated based on the method from Collyer et al,[51] for each phonation, the original microphone signal was first filtered in order to separate it into low-frequency (0 to 2000 Hz, low-pass filter) and high-frequency (2000 to 4000 Hz, band-pass filter) bands. The two resulting waveforms were then processed to obtain their corresponding sound level, using the "C" frequency weighting and the "special" time weighting, representing the sound levels in the frequency bands 0-2000 Hz (low frequency) and 2000-4000 Hz (high frequency). The spectral balance (SB) was calculated as the difference between the high-frequency sound level and the low-frequency sound level. One value of the SB was kept every 40 ms (1920 samples).

Difference between the levels of the first and the second harmonics (L_{H1}-L_{H2}) – this parameter was used to assess the relative importance of the first harmonic in the acoustic spectrum.[52,53] First, the signal was divided into 40-ms (1920 samples) segments with a 50% overlap between neighboring segments. After applying a Hanning window, the power spectrum of each segment was calculated from 100 Hz to 5000 Hz at 24 bands per octave, using the "poctave" Matlab built-in function. The power levels of the first (L_{H1}) and second (L_{H2}) harmonics were defined as the maximum value of the power spectrum in the frequency bands around the intended f_o (± 3.5 semitones) and twice the intended f_o (± 3.5 semitones), respectively. L_{H1}-L_{H2} was then calculated for each segment as the difference between the power level of the first harmonic and the power level of the second harmonic (in dB).

Electroglottographic parameter

EGG amplitude – this parameter was used as an indication of the relative vocal fold contact area during phonation. We expected it to be larger in chest-like than in head-like registers. For every phonation, we divided the EGG signal into 20-ms (960 samples) windows and calculated the RMS value of the signal in each of these windows.

Statistical analysis

We first verified whether the parameters were normally distributed using the Kologomorov-Smirnov test. Then, in order to find whether the differences between the chest-like and head-like register were statistically significant, we paired the mean chest-like register values with the corresponding mean head-like register values at each pitch, and we performed paired, one-sided statistical tests on the entire range, for each parameter. We used Student's t-tests on each of the analyzed parameters which were normally distributed, and Wilcoxon signed-rank tests on the other parameters. The significance level was set to $\alpha = 0.05$ (5%).

RESULTS

Listening tests

Figure 3 presents the results for the listening tests in the form of an "agreement score" summing the opinions of the listeners, as mentioned in the methods section (2.2) – the larger the score, the more frequently the listeners agreed that the perceived voice register was the same as intended by the singer. To account for a potential bias due to the familiarity of the "insiders" group with the recorded data, we separated the results of the "insiders" and "outsiders" groups. Figure 3 is divided into three panels showing the agreement scores of the insiders group (Figure 3A), the outsiders group (Figure 3B), and the total score regardless of the group (Figure 3C). Each panel was divided into an "agreement" zone (values between 0 and $+n$) and a "disagreement" zone (values between $-n$ and 0). Table 2 also presents the average results of each group, as well as the individual results from each listener.

Within the "insiders" group ($n = 6$ answers per pitch), the majority of the answers agreed with the singer's intent, thus correctly recognizing the head-like register having more head-like quality than the chest-like register (Figure 3A). Agreement was not reached for three tones: on C4, F5 and B5, the chest-like and head-like register phonations were not clearly distinguishable by these listeners. The agreement score was 0 for the C4 tone, indicating a "random" agreement (there was the same amount of "correct" and "incorrect" answers), and it was negative for F5 and B5,

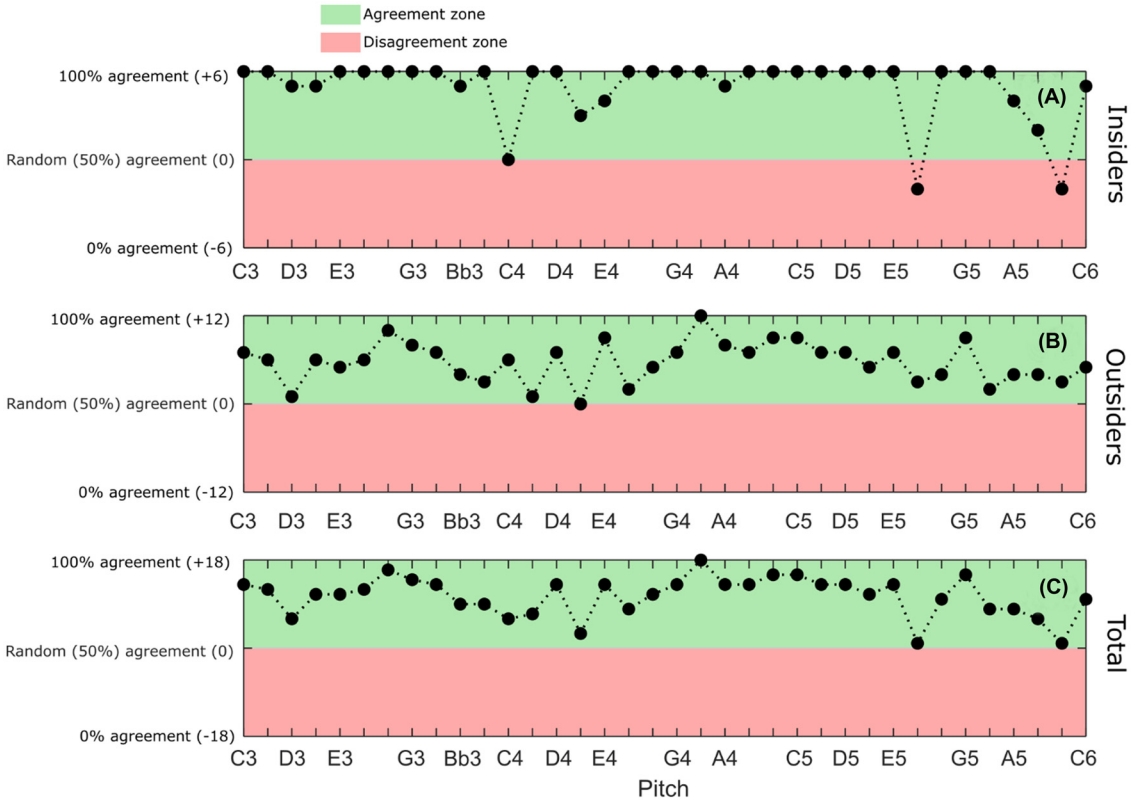


FIGURE 3. "Agreement score" from the listening tests. (A) insiders group: 6 tests; (B) outsiders group: 12 tests; (C) total: insiders + outsiders, 18 tests. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

indicating that there were more answers disagreeing with the singer's intent than answers agreeing with it, for those two tones. On average, the correct identification rate for the insider's listener group was 89% and the listeners were in 100% agreement with the singer's intention on 24 out of 36 pitches (see Table 2).

The listening test results for the "outsiders" group ($n = 12$ answers per pitch) are shown in Figure 3B. Generally, the agreement rate was considerably lower, i.e., 64 % (see Table 2). Out of 12 performed tests on each tone, only one tone (Ab4) had a 100 % correct identification rate, indicating that the listeners often did not find the register differences to be clear for the other tones. Nevertheless, for all except one pitch (Eb4, where the agreement score was 0), the majority of the answers agreed that the produced head-like register had more head-like quality than the produced chest-like register (agreement score > 0, as seen in Figure 3B). Compared with the results of the insiders, the pitches C4, F5 and B5 did not have a visibly lower difference score than the average.

TABLE 2.
Main Results From the Listening Tests

Listeners group		Individual percentage of agreement with the singer's intent	Average percentage of agreement with the singer's intent	Number of pitches with a 100 % agreement rate (out of 36 pitches)
Outsiders	O1	71	64	1
	O2	83		
	O3	39		
	O4	54		
	O5	51		
	O6	86		
Insiders	HL	90	89	24
	JGS	90		
	LP	88		
Total			73	1

TABLE 3.
Test-Retest Results of the Listening Tests for the Individual Raters and the Correlations Between the Results and the Listener Experience E_R From Table 1

Participant	Weighted Cohen's Kappa κ	Strength of agreement (see[54])	Number of identical answers between the two attempts N_I, out of 36 (percentage)
Outsiders (mean)	0.17		22 (61%)
O1	0.13	Slight	20 (56%)
O2	0.08	Slight	26 (72%)
O3	-0.06	Poor	11 (31%)
O4	0.19	Fair	18 (50%)
O5	0.40	Moderate	23 (64%)
O6	0.29	Fair	30 (86%)
Insiders (mean)	0.29		31 (86%)
HL	-0.08	Poor	29 (81%)
JGS	0.58	Moderate	33 (92%)
LP	0.36	Fair	30 (83%)
Parameters tested for correlation		(E_R, κ)	(E_R, N_I)
Pearson's correlation coefficients r(A, B)		0.23	0.45
Significance of the correlation p		0.55	0.21

Figure 3C displays the "total" results, which consists of the sum of the "insiders" and the "outsiders" groups. For these results, all tones were in the "agreement" zone, indicating that majority of the answers agreed with singer's intention, perceiving the head-like register having more head-like quality than the chest-like register.

In order to test the reliability of the listeners in their ratings, we investigated the intra-rater repeatability by calculating Cohen's weighted Kappa κ values for all the listeners (see Table 3). The values were between -0.08 and 0.58, indicating that the ratings were rather inconsistent. In only two raters the value of kappa was above 0.4, indicating a moderate repeatability (according to Landis and Koch[54]). The table also includes the number of identical answers between the two test attempts N_I for each participant, as another way of quantifying the repeatability with the values ranging between 31 and 92% agreement. Finally, the table lists the Pearson's correlation coefficient r estimating the degree of correlation between the self-rated listeners' experience in recognizing registers E_R (see Table 1), Cohen's Kappa, and the number of identical answers between the two listening test attempts. We observed that there was no correlation between the listeners' experience in recognizing registers and Cohen's Kappa (r = 0.23), and between the listeners' experience in recognizing registers and the number of identical answers between the two attempts (r = 0.45). None of the correlations was found significant at the 5% threshold.

As expected, the number of identical answers between the two test attempts was higher for the insiders group than for the outsiders group (except the participant O6). The outsiders group had an average of 22 identical answers between the two test attempts (61%), whereas the insiders group had an average of 31 identical answers between the two test attempts (86%). In general, we observed no clear correlation between this number and their self-rated ability to recognize the registers.

High-speed video analysis

Visual evaluation

At first, we evaluated the laryngeal configurations and vocal fold vibratory patterns from the high-speed video recordings visually. Figure 4 shows an overview of the pitch ranges where consistent visual differences were observed between the chest and head registers. For convenience, we defined R_{Lo}, R_{Mid} and R_{Hi} as the low (C3 to G4), middle (Ab4 to B4) and high (C5 to C6) ranges respectively.

The head-like register phonations exhibited a consistent decrease in the ventricular fold medialization for the lower pitch range, from C3 (131 Hz) to F4 (349 Hz), as indicated in Figure 4. The difference is shown on Figure 5 for the pitches Bb3 (233 Hz) and D4 (294 Hz). For each pitch, the left and middle panels show a high-speed video frame corresponding to the maximum glottal opening in the chest-like and head-like register respectively. On those frames, we drew the contour of the glottis and of the surrounding structures, such as the arytenoids, the aryepiglottic folds, the ventricular folds and the epiglottis. The right panel shows the sketches of the chest-like and head-like register contour overlaid for comparison. For the lower pitches Bb3 (first row) and D4 (second row), the ventricular folds were less adducted in the head-like register, which is indicated by solid black arrows. In the higher pitch range, ie, above F4 (349 Hz), differences in adduction were seen in some pitches but were not consistent. Furthermore, in that range it was not clear whether the differences were really reflecting a change in the laryngeal adjustment or were simply caused by a shift of the endoscope or a change of the larynx height.

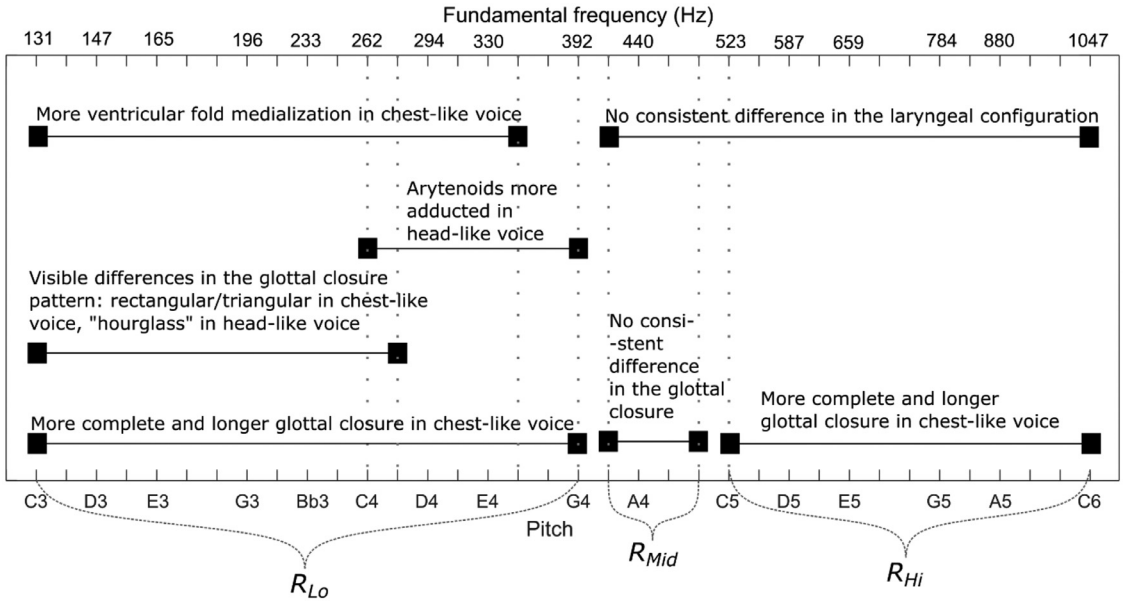


FIGURE 4. Diagram summarizing the visual differences observed between the chest-like and head-like phonations in the high-speed videos and the corresponding ranges where those differences appeared. R_{Lo}, R_{Mid} and R_{Hi} define the low (C3 − G4), middle (Ab4 − B4) and high (C5 − C6) ranges respectively. Examples of high-speed video recordings for the tones A3, E4, A4 and C6 are provided as supplementary videos S1A−S8A and the corresponding vocal sounds are provided as supplementary files S1B−S8B.

Figure 4 also indicates consistent differences in the arytenoids adduction in the range C4 (262 Hz) − G4 (392 Hz). Figure 5 shows those differences for the pitch D4 (294 Hz, second row), where the arytenoids were clearly more adducted in the head-like register (indicated by solid grey arrows). Finally, Figure 5 shows differences in the epiglottis position, where in the chest-like register the epiglottis was moved posteriorly (indicated by dashed black arrows). These differences were noticed visually on the pitches G3 (196 Hz), Bb3-Ab4 (233-415 Hz), and B4-Bb5 (494-932 Hz).

Consistent differences between the chest-like and head-like registers were observed in glottal closure between C3 (131 Hz) and G4 (392 Hz), and between C5 (523 Hz) and C6 (1047 Hz). Inside those pitch ranges, the glottal closure was fuller and longer in chest-like register. Examples of these differences are shown on Figure 6, which displays the juxtaposed high-speed video frames of one or two vocal fold oscillation cycles for several pitches, in both registers. For the pitch A3 (220 Hz), in chest-like voice the glottis fully closed along its membranous part, leaving a very small posterior gap between the arytenoids (Figure 6A). In contrast, in head-like register only a smaller portion of the glottis was closed (Figure 6B). In addition, the closure duration was visibly longer in chest-like than in head-like register. For the pitches Gb5 (740 Hz) and A5 (880 Hz), similarly to the lower pitches, the membranous portion of the glottis fully closed in chest-like register (Figure 6G and I), whereas an anterior gap was always visible in head-like voice (Figure 6H and J). In the middle of glottis, the closure also appeared to have a longer duration in chest-like than in head-like register.

Further, we observed clear differences in the vocal fold vibration patterns between C3 (131 Hz) and Db4 (277 Hz). In this pitch range, the chest-like phonations exhibited a "rectangular" vibration pattern, ie, the glottis closed and opened as a whole (as visible for the pitch A3 in Figure 6A), whereas the head-like phonations exhibited an "hourglass" glottal closure, ie, the glottis was closed only in its middle section, always leaving a small gap both anteriorly and posteriorly (Figure 6B).

Finally, in the pitch range from Ab4 (415) to B4 (494 Hz) the differences in the glottal closure were small, hardly noticeable visually, and rather inconsistent. For the pitch Ab4 (415 Hz), the glottis opening appeared slightly more abrupt in the chest-like register (Figure 6C) than in head-like register (Figure 6D), which is indicated in Figure 6C and D with red rectangles, but this could possibly be related also to slightly different image sampling moments within the vibratory cycle between the two phonations. The pitch A4 (440 Hz) showed very little but noticeable differences between the chest-like and head-like phonations, which are again shown by red rectangles: in the chest-like register (Figure 6E), the glottis opening started from the anterior section, whereas it started from the posterior section in the head-like register (Figure 6F).

Quantitative analysis of the segmented high-speed videos

Closed quotient (CQ_{PVG}). The results of the quantitative CQ_{PVG} analysis for the anterior, middle and posterior glottis are shown in Figure 7. The CQ_{PVG} values show a remarkable transition around the pitch Ab4 (415

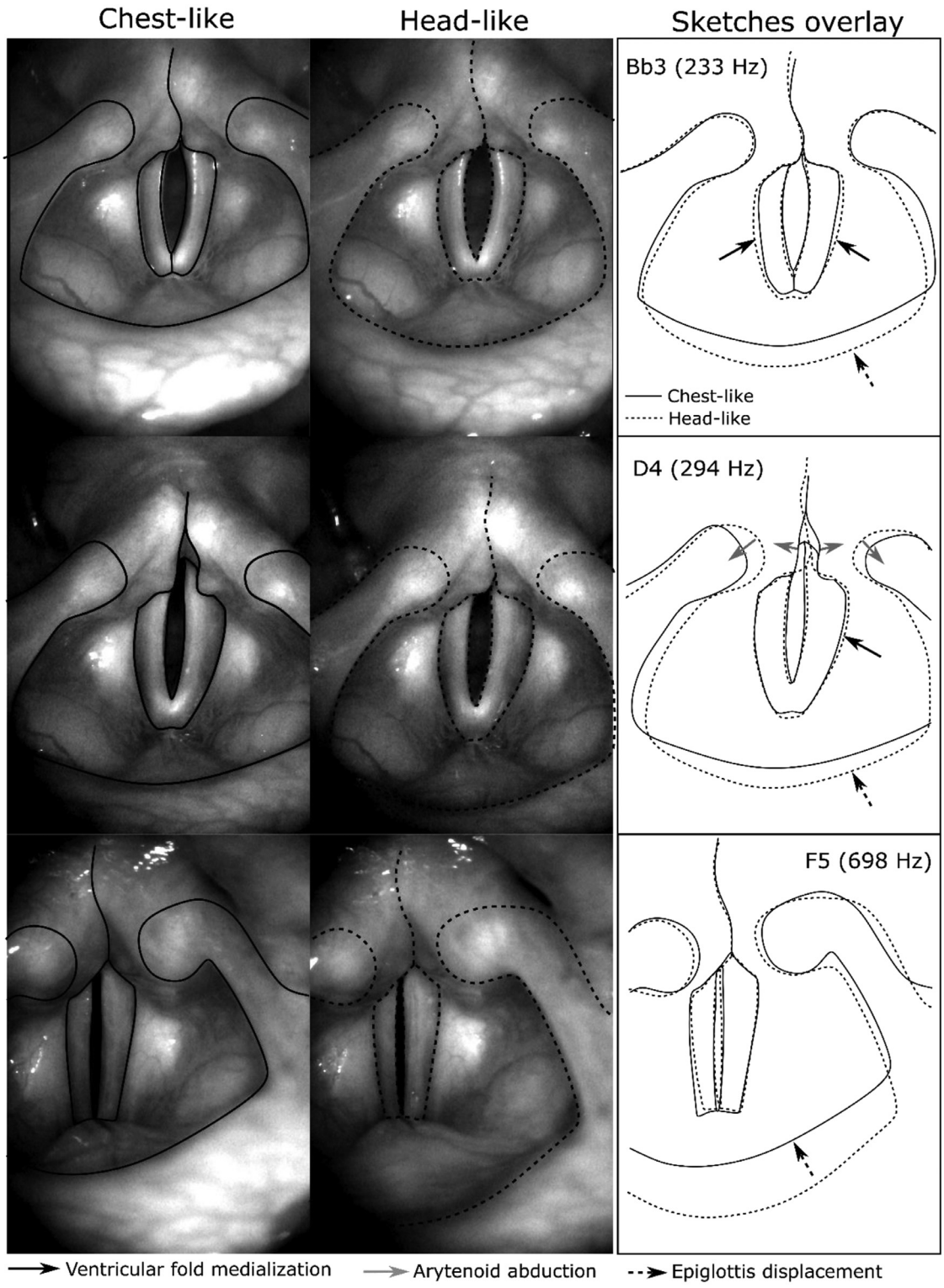


FIGURE 5. Frames of the high-speed videos and sketches of the laryngeal configuration at maximum glottal opening for the pitches Bb3, D4 and F5. The solid lines correspond to the chest-like register and the dashed lines to the head-like register. The arrows indicate recognizable differences between the two registers and the direction of change from the head to the chest register: the solid black arrows indicate a change in the ventricular fold medialization, the solid grey arrows indicate a change in the arytenoid adduction, and the dashed black arrows indicate a displacement of the epiglottis.

Hz) separating the lower pitches from the rest of the pitch range. In the lower pitch range (R_{Lo}), the differences between the chest-like and head-like registers were more apparent. Here, the CQ_{PVG} showed consistently larger values for chest-like than for head-like register at the middle and anterior locations along the glottis (see

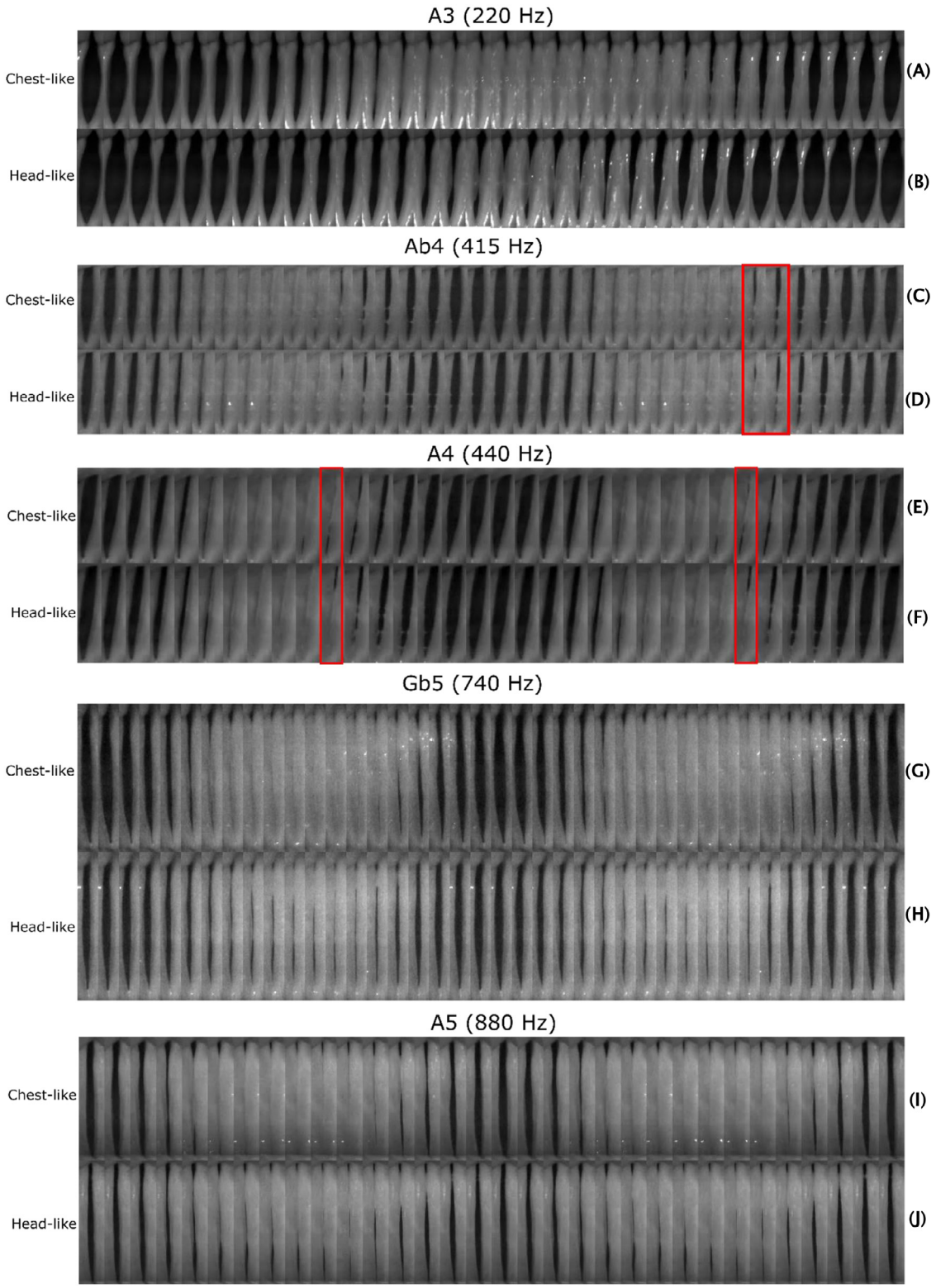


FIGURE 6. Frames from the high-speed videos comparing one or two glottal oscillation cycles for the chest and head registers for different pitches. The red rectangles emphasize the slight differences observed between the two registers for the pitches Ab4 (415 Hz) and A4 (440 Hz). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Figure 7A and B). At the posterior glottis, the CQ_{PVG} values were also generally larger for the chest-like register except for the pitches A3 (220 Hz), C4 (262 Hz) and Eb4 (311 Hz). These results agree with the visual differences in the vocal fold contact (incomplete and shorter glottal closure in the head-like register) observed in the high-speed video frames of the lower-pitched phonations (recall Figure 6).

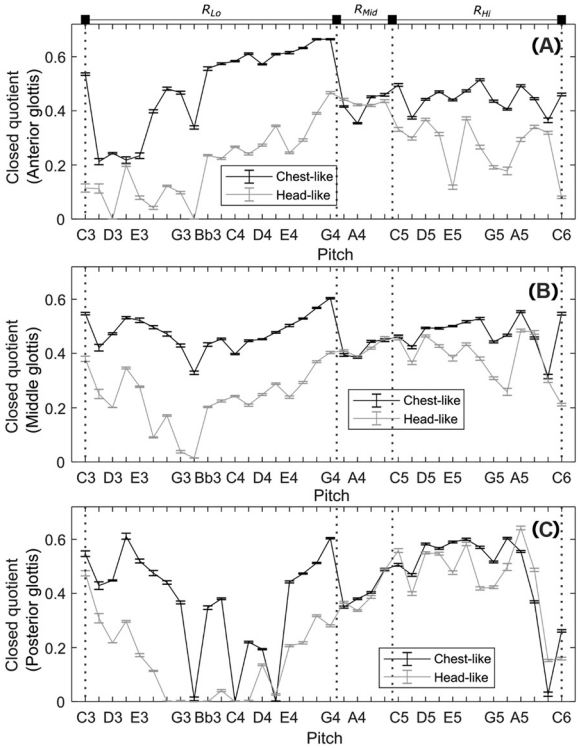


FIGURE 7. Median value of the cycle-averaged local closed quotients from the (A) anterior, (B) middle and (C) posterior thirds of the glottis. The error bars represent the standard error of the mean at the representative location for the anterior, middle and posterior glottis as determined by the median. The low, middle and high ranges are indicated by vertical dotted lines to highlight consistent differences. These ranges correspond to those indicated in Figure 4 and are also indicated in the following figures, for comparison purposes.

In the middle pitch range (R_{Mid}), around Ab4-B4 (415 - 494 Hz), the CQ_{PVG} values of the chest-like and head-like registers were similar, sometimes even unexpectedly slightly larger for the head-like register, for example at the pitch Ab4, reflecting the slightly shorter glottal closure duration observed in the chest-like register (see Figure 6C-D). In the high pitch range (R_{Hi}) from C5 to C6 (523 - 1047 Hz), at the anterior glottis, the CQ_{PVG} values were again larger for the chest-like register, even though the difference appeared smaller than in the low range below Ab4 (415 Hz) (Figure 7A). At the middle glottis, the CQ_{PVG} was higher in the chest-like register only between D5 (587 Hz) and A5 (880 Hz), and this difference appeared smaller than at the anterior glottis (Figure 7B). Finally, at the posterior glottis, the CQ_{PVG} differences between the chest-like and head-like registers were not consistent from C5 to C6 (Figure 7C). Again, these findings at high pitches agree with the visual observations of the high-speed video frames in Figure 6, where a small gap was observed in the head-like register at the anterior (and sometimes middle) glottis, but not in the posterior glottis (see pitches Gb5 and A5 in Figure 6H and Figure 6J). One-sided paired Student's t-tests showed that the CQ_{PVG} was significantly higher in chest-like than in head-like register at the anterior ($P < 0.001$), the middle ($P < 0.001$) and the posterior ($P < 0.001$) positions.

Normalized amplitude quotient (NAQ_{GA}) and closing quotient (CgQ_{kym}). In contrast to the CQ_{PVG}, which showed different behavior at the lower, middle and high pitch ranges, the behavior of NAQ_{GA} was found remarkably consistent across the entire pitch range. NAQ_{GA} was always lower in the chest-like register than in the head-like register (see Figure 8A). The smallest NAQ_{GA} difference was observed for the transition pitch Ab4 (415 Hz). The CgQ_{kym} showed similar trends to NAQ_{GA}, as expected: it was consistently lower in the chest-like than in the head-like register, indicating that the closing phase was shorter in the chest-like register, except for the pitch B5 (988 Hz), where the values were very close (Figure 8B). One-sided paired Student's t-tests showed that both parameters were significantly lower ($P < 0.001$) in the chest-like than in the head-like register.

Speed quotient (SQ_{kym}). The values of the SQ_{kym}, obtained from the segmented digital kymograms at the

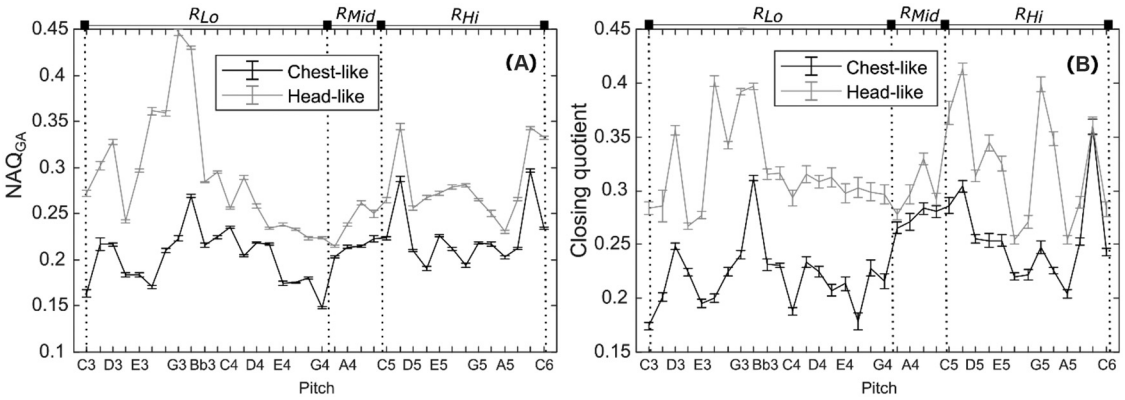


FIGURE 8. Mean Normalized Amplitude Quotient NAQ_{GA} (A) and Closing Quotient CgQ_{kym} (B) for every pitch. The error bars represent the standard error of the mean.

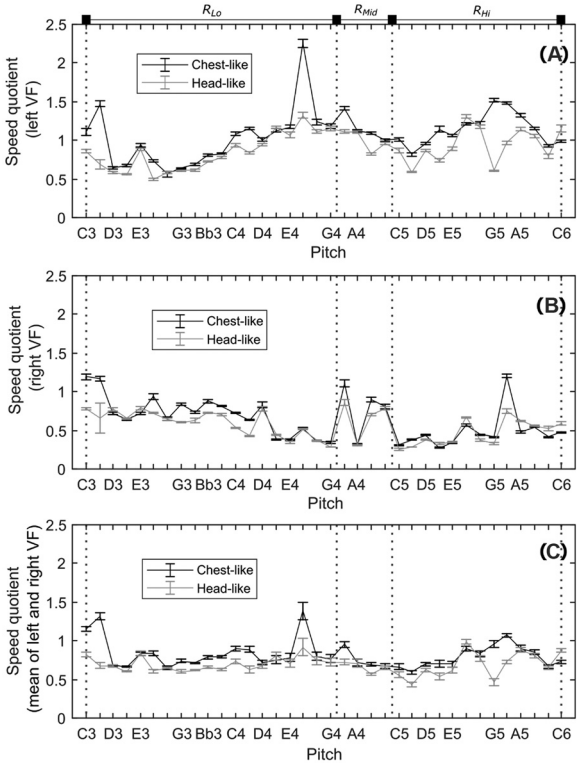


FIGURE 9. Mean speed quotient SQ_{kym} for the left vocal fold (A), right vocal fold (B) and the mean of the left and right vocal folds (C). The error bars represent the standard error of the mean.

place of maximum vibration amplitude in the middle of glottis, were rather different for the left and right vocal folds. It seemed the values followed an inverse trend for the left and right vocal fold as f_o increased: the values for the left vocal fold went up (Figure 9A), whereas the values for the right vocal fold went down (Figure 9B). The averages of the left and right values were mostly between 0.5 and 1 (Figure 9C) and did not exhibit any increasing or decreasing trend. Unlike the CQ_{PVG}, the SQ_{kym} values did not display a change in behavior across the low, middle and high pitch ranges. One-sided paired Student's t-tests showed that the SQ_{kym} had significantly higher values in chest-like register than in head-like register for the left ($P < 0.001$) and right ($P = 0.004$) vocal folds, as well as for the left-right average values ($P < 0.001$). The SQ_{kym} values did not allow clearly distinguishing between the chest-like and head-like phonations across all pitches, however.

Vertical phase differences (VPD_{kym}). The estimated vertical phase differences (VPD_{kym}) between the vibrations of the upper and lower vocal fold margins for the chest-like and head-like registers are shown in Figure 10. Similar to the CQ_{PVG} values, the VPD_{kym} parameter shows a remarkable transition around the note Ab4, separating the low pitch range from the rest. At the low pitch range, the VPD_{kym} values are notably larger for the chest-like register than for the head-like register, as hypothesized. From Ab4 up, the VPD_{kym} parameter could not reliably distinguish the two registers, however. Overall, however, the one-sided paired Student's t-test test showed that the VPD_{kym} was significantly higher ($P < 0.001$) in the chest-like than in the head-like register.

Acoustic analysis

Figure 11 shows the results from the acoustic analysis of the microphone signal. None of the analyzed acoustic parameters showed consistent differences for distinguishing the chest-like and head-like phonations across the whole pitch range. The SPL (Figure 11A) was consistently higher in chest-like register for the pitches from C3 (131 Hz) to Bb3 (233 Hz), and from Db5 (554 Hz) to C6 (1047 Hz). In this range, the differences in SPL ranged from about 0.9 dB (at C3) to slightly under 10 dB (at E5 = 659 Hz). In the middle pitch range from B3 (247 Hz) to C5 (523 Hz), there appeared to be no consistent differences in the SPL, but for two specific tones (A4 and B4) the SPL was about 2.5 dB higher in chest-like than in head-like register. The one-sided paired Student's t-test showed that the SPL was significantly higher in chest-like than in head-like register ($P < 0.001$).

The spectra of the chest-like and head-like phonations were visually quite similar. A more detailed analysis revealed that the head-like phonations (Figure 11C) had a slightly lower energy in the higher partials around 3000 Hz than the chest-like phonations (Figure 11B). The relative amplitude of the first harmonic was also generally higher in the head-like register. In both registers, the relative amplitude of the third harmonic abruptly decreased from Gb3 (185 Hz) to around C4 (262 Hz) - this is shown by the dashed white ellipses in Figure 11B and C. This spectral change did not show up to be reflected in the laryngeal configurations or in the physiologic and oscillatory parameters investigated here, however.

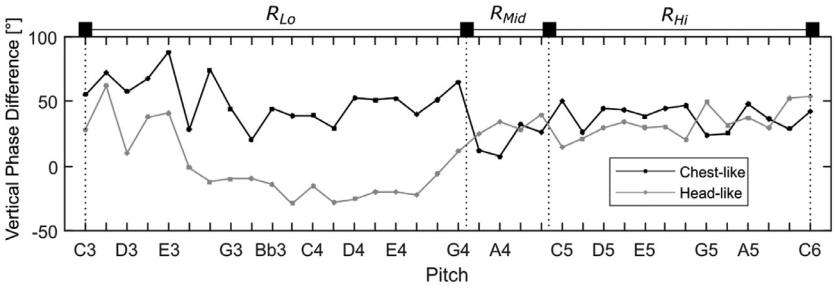


FIGURE 10. Estimated vertical phase differences between the lower and upper vocal fold margins (average value of the left and right vocal folds).

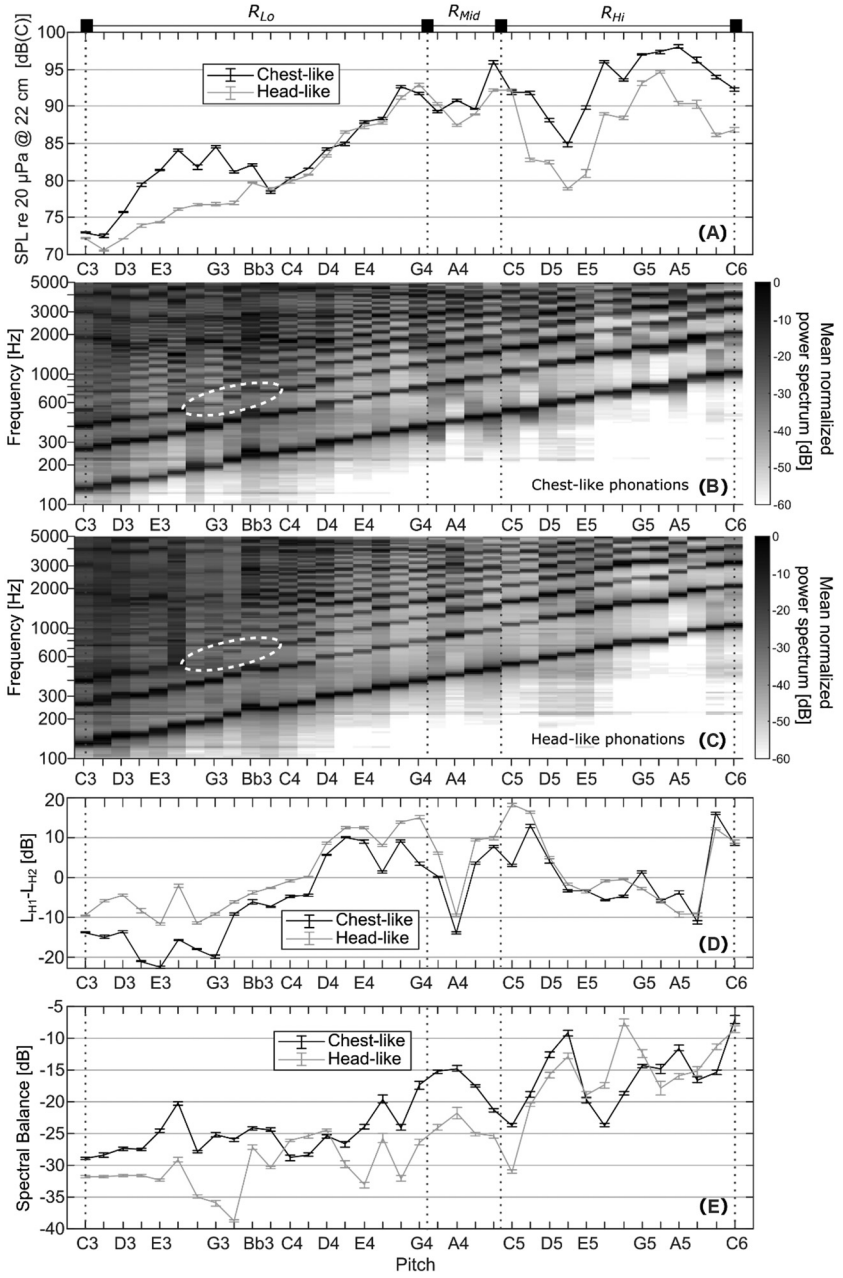


FIGURE 11. Results of the acoustic analysis of the microphone signal. (A) Mean sound pressure level (SPL). (B) Average power spectra for the chest-like register phonations. (C) Average power spectra for the head-like register phonations. (D) Mean difference between the levels of the first and second harmonic (L_{H1}-L_{H2}). (E) Mean spectral balance (difference of the sound level between the 0-2000 Hz and 2000-4000 Hz frequency bands). The dashed white ellipses in panels (B) and (C) indicate the abrupt decrease of the amplitude of the third harmonic. The error bars in panels (A), (D) and (E) represent the standard error of the mean.

The values of L_{H1}-L_{H2} were generally higher in the head-like register (Figure 11D), except for the pitch E5 (659 Hz), pitches from G5 (784 Hz) to A5 (880 Hz), and B5 (988 Hz). Higher values of L_{H1}-L_{H2} indicate that the first harmonic contains more energy relative to the second harmonic. A one-sided paired Student's t-test indicated that the L_{H1}-L_{H2} values were significantly higher in the head-like register ($P < 0.001$).

The values of the spectral balance (Figure 11D) were mostly higher in the chest-like phonations, but this did not hold for the pitches from C4 (262 Hz) to D4 (294 Hz), from E5 (659 Hz) to G5 (784 Hz), and from Bb5 (932 Hz) to B5 (988 Hz). Higher values of the spectral balance indicate more energy in the high-frequency band compared to the low-frequency band. A one-sided paired Student's t-test revealed that the spectral balance was significantly higher in chest-like register ($P < 0.001$).

Amplitude of the EGG signal

Amplitude of the EGG signal was analyzed to compare the relative vocal fold contact area between the two registers. The mean RMS amplitude was almost always higher in the

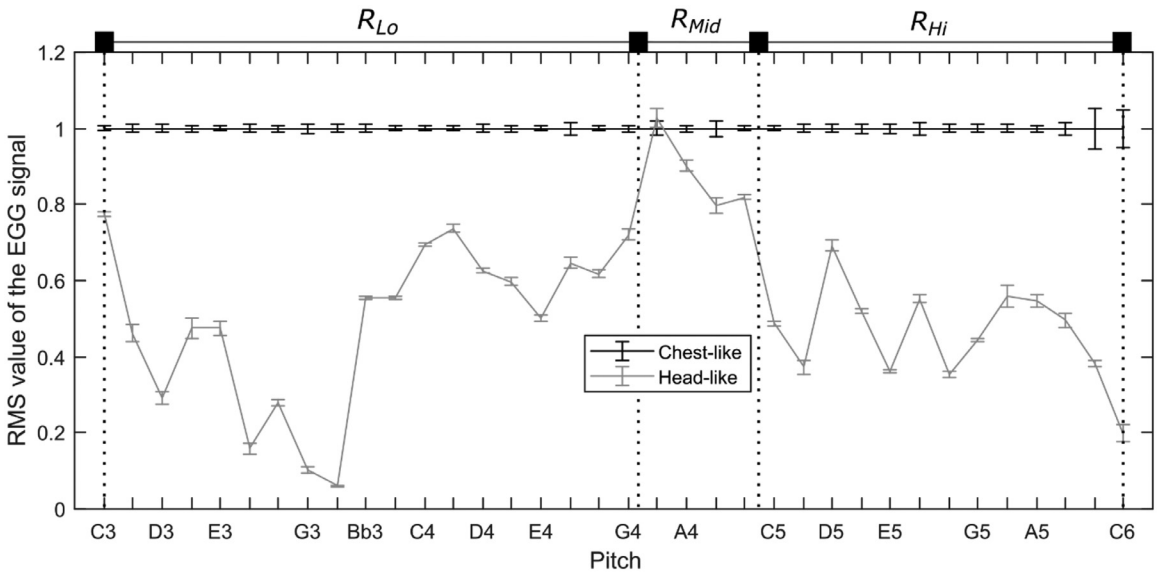


FIGURE 12. Mean RMS value of the EGG signal. The error bars represent the standard error of the mean. For each pitch, the mean value and the standard error were normalized by the mean value of the chest-like register.

chest-like register than in the head-like register (see Figure 12), as hypothesized. The only pitch where the head-like register value was slightly higher than the chest-like register value was Ab4 (415 Hz). Statistically, the RMS amplitude of the EGG was overall significantly higher in the chest-like register ($P < 0.001$), as expected.

DISCUSSION

As we mentioned in the introduction, both chest and head registers are usually assumed to be limited in pitch, suggesting that the chest register cannot be used above a certain pitch and the head register cannot be used below a certain pitch, even though the limits can vary to some extent between individuals. Consequently, being able to freely choose either register regardless of the pitch throughout a singing range spanning over three octaves is a rather unusual claim. Nevertheless, our objective analysis of the vocal fold vibratory patterns using laryngeal high-speed videoendoscopy and electroglottography confirmed the claim of the singer by revealing consistent differences between the two registers produced. Most consistently, the normalized amplitude quotient (NAQ_{GA}, obtained from the GAW) and the closing quotient (CgQ_{kym}, obtained from the digital kymograms) showed distinctly lower values in the chest-like register at every pitch (except one pitch for the CgQ_{kym}), and the RMS amplitude of the EGG signal showed consistently higher values in the chest-like register at every pitch except one. However, there were also remarkable differences in the laryngeal configuration separating the low-, middle-, and high-pitched phonation regions (Figure 4). These indicate that our professional singer employed subtle laryngeal control mechanisms for the chest-like and head-like phonations on top of the traditionally recognized low-pitched chest versus high-pitched head register phenomena.

Perceptually, while the listeners agreed with the singer's intention in the majority of the cases, distinguishing between the two registers showed up to be a challenging task. The two registers were recognizable by the "outsiders" listeners, on average, only in 64% of the cases (see Table 2). This suggests that the listeners had different expectations about the registers, even those who were singing teachers with many years of experience. Some listeners might be used to focus on articulation differences when recognizing registers, and therefore the articulation limitations imposed by the endoscope insertion might have decreased their confidence when rating the samples. The "insiders" performed better, as expected, agreeing with the singer's intent in 89 % of the cases, on average. Among the "insiders", interestingly, the singer LP did not score as the best rater in the listening tests, showing only 83% test-retest agreement (recall Table 3) and correctly identifying her intended registers in 88% of the pitches, whereas the two other authors both correctly identified the registers in 90% of the pitches (recall Table 2). This is likely related to the fact that the singer uses also kinesthetic and bone-conduction-sound feedback when producing the registers and her auditory sensations were insufficient to correctly recognize the intended voice register in all the cases. This seemed to eliminate the potential bias of the singer as a listener judge; therefore, we kept her results as part of the insiders group.

Physiologically, our hypothesis (1) expected a longer and more pronounced contact in the chest-like register than in the head-like register. The values of the high-speed-video-based CQ_{PVG} and of the EGG amplitude generally supported this hypothesis at most pitches but not all of them. In the low (up to G4 = 392 Hz) and high (from C5 = 523 Hz to C6 = 1047 Hz) ranges, the CQ_{PVG} was always higher in the chest-like register at the anterior and middle glottis, suggesting a higher membranous vocal fold adduction. However, the values from the middle range (Ab4 = 415 Hz to

B4 = 494 Hz) showed much smaller differences, and were even higher in the head-like register for some pitches. The RMS amplitude of the EGG signal was also generally higher in the chest-like register and the differences were smaller in the middle range. Like the larger CQ_{PVG} value, a higher RMS amplitude of the EGG signal also indicates a stronger and longer relative vocal fold contact, but some role could be played here also by the different larynx height in the chest-like and head-like registers which could not be ruled out in this study. In principle, the EGG contact quotient could also have been used for estimating the relative duration of the glottal closure (see Herbst et al,[55]) but due to deviations of the EGG waveform shape from the commonly-observed shapes in some of our phonations we have not explored this parameter. The obtained results, nevertheless, provide strong evidence of a more membranous adduction in the chest-like register at most pitches, which supports the hypothesis of a stronger TA muscle activity in this register.

Hypothesis (2) expected the vertical phase differences (VPD_{kym}) between the vibrations of the lower and upper margins of the vocal folds to be larger in the chest-like register than in the head-like register. This was found true only in the low range (from C3 = 131 Hz to G4 = 392 Hz). It suggests that, for the higher pitches, either there was no substantial difference in the vocal fold thickness allowing larger VPD_{kym}, or that the VPD_{kym} is not the most robust parameter to detect the register differences there.

Hypothesis (3) expected the glottal closing speed to be higher in the chest-like register than in the head-like register. The observed NAQ_{GA} and CgQ_{kym} values support this hypothesis. The consistently lower NAQ_{GA} values indicate that the maximum closing speed was higher in the chest-like register throughout the entire pitch range. Similarly, the CgQ_{kym} values were consistently lower in the chest-like register, indicating that the relative duration of the closing phase was shorter than in the head-like register. Since the maximum speed of glottal closure and the relative duration of the closing phase were theoretically shown to be related to the vibration amplitude of the lower vocal fold margins,[30] those results suggest that the amplitude of the lower margin of the vocal folds was larger in the chest-like register. In addition, the analysis of the SQ_{kym} revealed that in the chest-like register the duration of the closing phase was generally shorter in relation to the duration of the opening phase, but these differences were less consistent across the entire pitch range than those found for the NAQ_{GA} and CgQ_{kym} values.

Hypothesis (4) expected more high-frequency energy in the sound spectrum of the chest-like compared to the head-like register. The results from the acoustic analysis of the microphone signal generally support this hypothesis (Figure 11D-E), but, again, not at all pitches. The increase in the energy of the higher partials could be caused by an increase in the vocal fold adduction and in the glottal closure duration,[25] which agrees with the findings from the analysis of the high-speed video recordings. Chest-like

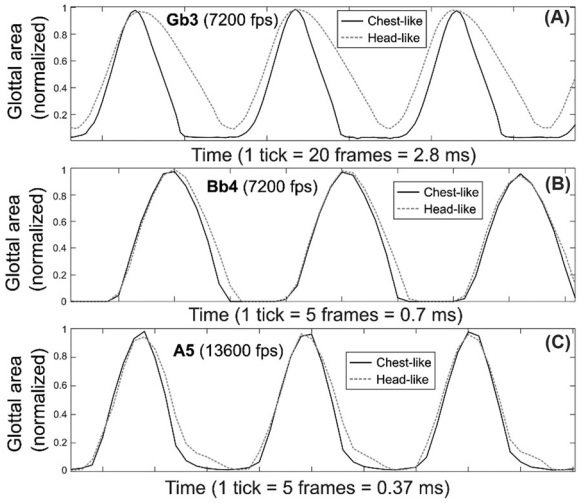


FIGURE 13. Glottal area waveforms (GAWs) for the chest-like and head-like register, representative of the low, middle and high pitch ranges shown in Figure 4: (A) GAW at the pitch Gb3 (185 Hz, R_{Lo}), (B) Bb4 (466 Hz, R_{Mid}), (C) A5 (880 Hz, R_{Hi}). Each waveform amplitude was normalized to 1. Notice the steeper declining slope in the GAWs of the chest-like registers.

register phonations were found to be produced mostly with higher SPL than the head-like phonations, although also not at all pitches (Figure 11A). Higher values of the SPL are generally associated with a higher phonatory effort, higher subglottal pressure,[56,57] and an increase in the maximum flow declination rate.[44,58] For technical reasons the subglottal pressure and glottal flow could not be measured in our study. The singer aimed at performing both registers with similar phonatory effort and therefore the differences in subglottal pressure were supposedly small; nevertheless, the influence of this control parameter cannot be totally neglected and should be addressed in future studies. The observed changes in the laryngeal configuration indicate, however, that the higher SPL values in the chest register observed here were not purely the result of respiratory factors but there was a considerable influence of different laryngeal adjustments.

Overall, our results suggest that the glottal closing speed is one of the most relevant parameters for discriminating between chest-like and head-like registers across the entire pitch range. This can be derived from the values of the NAQ_{GA} and the CgQ_{kym} (see Figure 8), which were both consistently lower (except for one pitch in the CgQ_{kym}) in the chest-like register, indicating shorter closing duration and thus higher closing speed of the vibrating vocal folds. More directly, the faster closure can be observed on the GAW which is more skewed to the right in the chest-like register (see Figure 13). As the closing movement is expected to depend mainly on the vibration of the lower margin of the vocal folds,[30,59] higher values of the closing speed indicate faster movement and a larger oscillatory amplitude of the lower margin of the vocal folds in chest-like register. Physiologically, this could be achieved by an increased TA

muscle activity bulging the vocal folds and pushing the lower part of the glottis in the medial direction so that the lower-margin oscillations could be more strongly excited by intraglottal aerodynamic forces. This mechanism would be in line with previous EMG studies showing more TA activity in chest than in head register.[6-9] Some role in the GAW skewing could also be played by the source-filter interaction (ie, the interactions of the oscillating glottis with the vocal tract acoustics). Despite of the rigid laryngoscope limiting the vocal tract adjustments, our laryngeal images revealed slight changes of the lower vocal tract geometry by varying the epiglottis and arytenoids position between the two registers (recall Figure 5). More research is needed to elucidate the degree of possible influence of the vocal fold bulging and of the source-filter interaction on the faster glottal closing speed observed here, however. Remarkably, the recent study of Patel et al. (2022)[60] also found the closing speed and closing duration to be among the most important factors for differentiating specific types of phonations.

Interestingly, within the range C4 − G4 our professional singer abducted the arytenoid cartilages in the chest-like register while pressing the arytenoids together in the head-like register (recall Figure 5, pitch D4). This contrasts with the results in an untrained female singer published by Švec, Sundberg and Hertegård[14] who reported more adducted arytenoid cartilages in chest than in head register. This discrepancy could possibly be related to the differences between naïve and highly skilled professional singers. While in the case of the naïve untrained singer the release of the arytenoid adduction was related to an abrupt transition from low-pitched chest to high-pitched head register,[14] in our professional singer no such abrupt transition was identified. The skilled singers usually aim at equalizing the voice quality throughout the entire pitch range; controlling the arytenoid adduction independently from the glottal adduction can serve as a possible means of achieving the goal. In this perspective, it appears important to distinguish between the cartilaginous and membranous glottal adduction when investigating the subtle laryngeal mechanisms of voice quality and register control in singing.[28,29,61]

Besides analyzing the separate register phonations of our singer, we also analyzed the phonations when the singer was switching back and forth between the two registers at the same pitch. We have not found any clear register breaks when switching from one register to another, which contrasts with Mal Webb's clearly audible breaks occurring at every change of register.[23] This difference might be related to the generally higher extent of pitch jumps in male than in female singers (see e.g.,[62,63]), suggesting that female singers generally have a smoother transition between registers. Alternatively, the discrepancy could be due to differences in both singers' training, as well as intentions to try smoothing (or not) the register transitions.

It is unknown how gender might influence the ability to freely choose either chest-like or head-like register regardless of the pitch. When experimenting with her students, our singer has informally noticed that female singers have little problem producing the chest-like and head-like sounds on all notes of their range, whereas this ability seems less common with male singers. Further work has to be done to find out possible gender-related differences.

It is worth noting that this study was performed only with one singer and our results are not a reflection of the general case. Nevertheless, the ability of the singer to shift between chest-like and head-like registers across a large range of pitches allows investigating which parameters appear the most consistent for differentiating the chest and head registers in highly skilled singers. This provides highly useful insights as such singers are expected to have enhanced physiological control of the laryngeal configuration surpassing that of naïve singers.

CONCLUSION

This case study offers a deeper insight into the laryngeal adjustments utilized by a female professional singer for changing the voice register. Our results are unique due to the fact that the singer showed an ability to use chest-like and head-like registers over a very large pitch range of three octaves. The most consistent parameter differing the two registers across the whole pitch range was NAQ_{GA} (from the glottal area waveform) showing always smaller values for chest-like than head-like register, and the closing quotient CgQ_{kym} (from the kymogram) revealing the same differences at all except one pitch. Biomechanically, these findings can be attributed to faster glottal closure and larger oscillatory amplitude of the lower margin of the vocal folds.

Additionally, the results showed longer closed phase at anterior glottis and larger EGG amplitudes in chest-like register at most pitches. These findings are in line with the previous observations that the chest register is mostly, but not necessarily, produced with more bulged vocal folds and more adducted membranous glottis, usually achievable through higher activity of the thyroarytenoid muscle. However, besides these rather consistent differences across the entire pitch range, we also noticed laryngeal, physiological and acoustic differences separating the low-, middle-, and high-pitched phonation regions indicating that the singer employed subtle laryngeal control mechanisms for the chest-like and head-like phonations on top of the traditionally recognized low-pitched chest and high-pitched head register phenomena. Acoustically, the register changes were not easily perceivable by listeners, however, which is in line with the long-term controversial opinions on registers based mainly on auditory perceptual evaluation.

AUTHOR CONTRIBUTIONS

H.L.: formal analysis, investigation, data curation, writing—original draft, visualization. L.P.: conceptualization, data acquisition, writing—review and editing. J.G.Š.: conceptualization, methodology, data acquisition, resources, writing—review and editing, supervision, funding acquisition.

ACKNOWLEDGMENT

The authors would like to thank the six listeners for their participation in the listening tests. A preliminary version of this work was presented at the ICVPB 2020, International Conference on Voice Physiology and Biomechanics in Grenoble, France on December 2-4, 2020 (online conference).

SUPPLEMENTARY DATA

Supplementary data related to this article can be found online at doi:10.1016/j.jvoice.2022.02.014.

Supporting information

Supplementary video S1A: High-speed video recording (excerpt) at the pitch A3 (220 Hz) in the chest-like register. Frame rate: 7200 fps. Resolution: 384*576 px.

Supplementary audio S1B: Vocal sound corresponding to S1A (pitch A3, 220 Hz, chest-like register). Sampling rate: 48 kHz, sample format: 24 bits.

Supplementary video S2A: High-speed video recording (excerpt) at the pitch A3 (220 Hz) in the head-like register. Frame rate: 7200 fps. Resolution: 384*576 px.

Supplementary audio S2B: Vocal sound corresponding to S2A (pitch A3, 220 Hz, head-like register). Sampling rate: 48 kHz, sample format: 24 bits.

Supplementary video S3A: High-speed video recording (excerpt) at the pitch E4 (330 Hz) in the chest-like register. Frame rate: 7200 fps. Resolution: 384*576 px.

Supplementary audio S3B: Vocal sound corresponding to S3A (pitch E4, 330 Hz, chest-like register). Sampling rate: 48 kHz, sample format: 24 bits.

Supplementary video S4A: High-speed video recording (excerpt) at the pitch E4 (330 Hz) in the head-like register. Frame rate: 7200 fps. Resolution: 384*576 px.

Supplementary audio S4B: Vocal sound corresponding to S4A (pitch E4, 330 Hz, head-like register). Sampling rate: 48 kHz, sample format: 24 bits.

Supplementary video S5A: High-speed video recording (excerpt) at the pitch A4 (440 Hz) in the chest-like register. Frame rate: 7200 fps. Resolution: 384*576 px.

Supplementary audio S5B: Vocal sound corresponding to S5A (pitch A4, 440 Hz, chest-like register). Sampling rate: 48 kHz, sample format: 24 bits.

Supplementary video S6A: High-speed video recording (excerpt) at the pitch A4 (440 Hz) in the head-like register. Frame rate: 7200 fps. Resolution: 384*576 px.

Supplementary audio S6B: Vocal sound corresponding to S6A (pitch A4, 440 Hz, head-like register). Sampling rate: 48 kHz, sample format: 24 bits.

Supplementary video S7A: High-speed video recording (excerpt) at the pitch C6 (1047 Hz) in the chest-like register. Frame rate: 13600 fps. Resolution: 384*576 px.

Supplementary audio S7B: Vocal sound corresponding to S7A (pitch C6, 1047 Hz, chest-like register). Sampling rate: 48 kHz, sample format: 24 bits.

Supplementary video S8A: High-speed video recording (excerpt) at the pitch C6 (1047 Hz) in the head-like register. Frame rate: 13600 fps. Resolution: 384*576 px.

Supplementary audio S8B: Vocal sound corresponding to S8A (pitch C6, 1047 Hz, head-like register). Sampling rate: 48 kHz, sample format: 24 bits.

Appendix

Instructions for the listening test

You will be provided with a file, which contains voice sounds of a singer. The singer produced sounds at different pitches across 3 octaves. The singer was able to produce the chest (modal) register, the head (flutey, loft) register, as well as produce the mix of these registers. Your task will be to rate the sound quality on the continuum between chest-like and head-like voice quality based on your own impressions.

1. Download the Visor software from Svante Granqvist's website (http://www.tolvan.com/index.php?page=/visor/visor.php)
2. Run the software (you will most likely need a Windows computer, I do not think MAC is supported...) as administrator and open the test.visor file I have sent you by email.
3. Click on "Start test!" and enter your name.
4. You should see a screen with many small icons looking like presents. Every one of these icons is the microphone signal of one phonation from the experiments, ranging from C3 to C6.
5. You can play each of them by right clicking on the present icons, as many times as you need.
6. The goal is to "rate" these phonations acoustically, whether they sound more like a chest voice or more like a head/flutey voice. To do this, drag the icons more to the bottom if you feel they sound more 'chest-like', or more to the top if you feel they sound more 'head-like'. It does not have to be always 100% chest (all the way to the bottom) or 100% head (all the way to the top) but it can also be somewhere in between. The scale is continuous.
7. You can freely arrange the icons according to pitch, from left to right if you find it more convenient. The horizontal position has no impact on the test results.
8. You have to do the test fully in one go, which means you cannot save your progress in the middle and continue later. Be aware that the test can take probably 10-30 minutes, maybe more.
9. Once you are done, please check all the icon sounds again by right clicking them and verify they are in the right vertical position according to your rating.
10. Finally, click on "Stop test!". The software will tell you where the result file ('Test.txt') will be saved. Please send me this file at hugo.lehoux.fr@gmail.com.
11. Please answer these questions, it will help us for the study.

- Are you a singer?
- If yes, for how many years have you been trained as a singer? (including singing in choir)
- How experienced are you in producing the chest (modal) register and the head (flutey, loft) register? (rate on a scale between 0 and 10, where 0 is no experience and 10 is an expert)
- How experienced are you in recognizing the chest (modal) register and the head (flutey, loft) register? (rate on a scale between 0 and 10, where 0 is no experience and 10 is an expert)
- How difficult was it to identify the register across the different pitches in this singer?
- Did you experience any problems with this study?
- Any other comment that you might find relevant.

Thank you! Your help is greatly appreciated.

REFERENCES

1. Stark JA. *BelCanto: A History of Vocal Pedagogy*. Toronto/Buffalo/London: University of Toronto Press; 1999.
2. Mörner M, Fransson F, Fant G. Voice register terminology and standard pitch. *STL-QPSR*. 1963;4:17–23.
3. Švec JG, Schutte HK, Miller DG. On pitch jumps between chest and falsetto registers in voice: data from living and excised human larynges. *J Acoust Soc Am*. 1999;106:1523–1531.
4. Hollien H. On vocal registers. *J Phon*. 1974;2:125–143.
5. Roubeau B, Henrich N, Castellengo M. Laryngeal vibratory mechanisms: the notion of vocal register revisited. *J Voice*. 2009;23:425–438.
6. Hirano M, Vennard W, Ohala J. Regulation of register, pitch and intensity of voice. An electromyographic investigation of intrinsic laryngeal muscles. *Folia Phoniatr Logop*. 1970:22.
7. Baer T, Gay T, Niimi S. Control of fundamental frequency, intensity and register of phonation. *Haskins Laborat Status Rep Speech Res*. 1976;45:175–185.
8. Kochis-Jennings KA, Finnegan EM, Hoffman HT, et al. Laryngeal muscle activity and vocal fold adduction during chest, chestmix, headmix, and head registers in females. *J Voice*. 2012;26:182–193.
9. Kochis-Jennings KA, Finnegan EM, Hoffman HT, et al. Cricothyroid muscle and thyroarytenoid muscle dominance in vocal register control: preliminary results. *J Voice*. 2014;28:652.e621–652.e629.
10. Henrich N, Roubeau B, Castellengo M. On the use of electroglottography for characterisation of the laryngeal mechanisms. In: *Proceedings of SMAC 2003, Stockholm Music Acoustics Conference*. 2, Stockholm, Sweden: KTH Royal Institute of Technology; 2003:455–458.
11. Roubeau B, Chevrie-Muller C, Arabia-Guidet C. Electroglottographic study of the changes of voice registers. *Folia Phoniatr Logop*. 1987;39:280–289.
12. Lindestad P-Å, Södersten M. Laryngeal and pharyngeal behavior in countertenor and baritone singing—a videofiberscopic study. *J Voice*. 1988;2:132–139.
13. Murry T, Xu JJ, Woodson GE. Glottal configuration associated with fundamental frequency and vocal register. *J Voice*. 1998;12:44–49.
14. Švec JG, Sundberg J, Hertegard S. Three registers in an untrained female singer analyzed by videokymography, strobolaryngoscopy and sound spectrography. *J Acoust Soc Am*. 2008;123:347–353.
15. Echternach M, Burk F, Köberlein M, et al. The influence of vowels on vocal fold dynamics in the tenor's passaggio. *J Voice*. 2017;31:424–429.
16. Echternach M, Burk F, Köberlein M, et al. Oscillatory characteristics of the vocal folds across the tenor passaggio. *J Voice*. 2017;31:381.e5–381.e14.
17. Echternach M, Burk F, Köberlein M, et al. Laryngeal evidence for the first and second passaggio in professionally trained sopranos. *PLoS One*. 2017;12: e0175865.
18. Echternach M, Dippold S, Richter B. High-speed imaging using rigid laryngoscopy for the analysis of register transitions in professional operatic tenors. *Logoped Phoniatr Vocol*. 2016;41:1–8.
19. Echternach M, Dippold S, Sundberg J, et al. High-speed imaging and electroglottography measurements of the open quotient in untrained male voices' register transitions. *J Voice*. 2010;24:644–650.
20. Kumada M, Kobayashi N, Hirose H, et al. Analysis of vocal fold vibration during register change by high speed digital imaging system. *Forum Acusticum Sevilla 2002 (CD-ROM)*. Madrid: Sociedad Española de Acústica; 2002.
21. Rubin HJ, Hirt CC. The falsetto. A high-speed cinematographic study. *Laryngoscope*. 1960;70:1305–1324.
22. Zañartu M, Mehta DD, Ho JC, et al. Observation and analysis of in vivo vocal fold tissue instabilities produced by nonlinear source-filter coupling: a case study. *J Acoust Soc Am*. 2011;129:326–339.
23. Webb M. Mal Webb Sideways Yodeling. https://www.youtube.com/watch?v=8_6nNWX7TTI (accessed 2021).
24. Colton RH. Spectral characteristics of the modal and falsetto registers. *Folia Phoniatr Logop*. 1972;24:337–344.
25. Titze IR.*Principles of Voice Production (second printing)*: national center for voice and speech iowa city IA 52242; 2000.
26. Hirano M. Morphological structure of the vocal cord as a vibrator and its variations. *Folia Phoniatr Logop*. 1974;26:89–94.
27. Hirano M. Phonosurgery: basic and clinical investigations. *Otologia (Fukuoka)*. 1975;21(suppl. 1):239–242.
28. Herbst CT, Qiu Q, Schutte HK, et al. Membranous and cartilaginous vocal fold adduction in singing. *J Acoust Soc Am*. 2011;129:2253–2262.
29. Herbst CT, Švec JG. Adjustment of glottal configurations in singing. *J Sing*. 2014;70:301–308.
30. Titze IR. Theoretical analysis of maximum flow declination rate versus maximum area declination rate in phonation. *J Speech Lang Hear Res*. 2006;49:439–447.
31. The Audacity Team. Audacity: free audio editor and recorder. https://www.audacityteam.org/(accessed 2021).
32. Granqvist S. The visual sort and rate method for perceptual evaluation in listening tests. *Logoped Phoniatr Vocol*. 2003;28:109–116.
33. Granqvist S. Visor: visual sort and rate. http://www.tolvan.com/index.php?page=/visor/visor.php (accessed 2020).
34. Herbst CT. CogBioCreateKymogram - A FIJI (ImageJ) plugin for creating videokymograms from high-speed movie data. http://www.christian-herbst.org/index.php?page=fiji (accessed 2021).
35. Novozámský A, Sedlář J, Zita A, et al. Image analysis of videokymographic data. *2015 IEEE International Conference on Image Processing*. (ICIP): IEEE; 2015:78–82.
36. Gómez P, Kist AM, Schlegel P, et al. BAGLS, a multihospital benchmark for automatic glottis segmentation. *Sci Data*. 2020;7:1–12.
37. Maryn Y, Verguts M, Demarsin H, et al. Intersegmenter variability in high-speed laryngoscopy-based glottal area waveform measures. *Laryngoscope*. 2020;130:E654–E661.
38. Kist AM, Gómez P, Dubrovskiy D, et al. A deep learning enhanced novel software tool for laryngeal dynamics analysis. *J Speech Lang Hear Res*. 2021;64:1889–1903.
39. Lohscheller J, Eysholdt U. Phonovibrogram visualization of entire vocal fold dynamics. *Laryngoscope*. 2008;118:753–758.
40. Švec JG, Granqvist S. Tutorial and guidelines on measurement of sound pressure level in voice and speech. *J Speech Lang Hear Res*. 2018;61:441–461.
41. Granqvist S, Švec JG. CalibrateVoiceSPL. Matlab scripts. http://www.mathworks.com/matlabcentral/fileexchange/64231-calibratevoicespl (accessed June 2, 2020).
42. Sielska-Badurek EM, Jędra K, Sobol M, et al. Laryngeal stroboscopy—Normative values for amplitude, open quotient, asymmetry

and phase difference in young adults. *Clin Otolaryngol*. 2019;44:158–165.
43. Alku P, Bäckström T, Vilkman E. Normalized amplitude quotient for parametrization of the glottal flow. *J Acoust Soc Am*. 2002;112:701–710.
44. Vilkman E, Alku P, Vintturi J. Dynamic extremes of voice in the light of time domain parameters extracted from the amplitude features of glottal flow and its derivative. *Folia Phoniatr Logop*. 2002;54:144–157.
45. Jiang JJ, Chang CI, Raviv JR, et al. Quantitative study of mucosal wave via videokymography in canine larynges. *Laryngoscope*. 2000;110:1567–1573.
46. Jiang JJ, Titze IR. A methodological study of hemilaryngeal phonation. *Laryngoscope*. 1993;103:872–882.
47. Jiang JJ, Zhang Y, Kelly MP, et al. An automatic method to quantify mucosal waves via videokymography. *Laryngoscope*. 2008;118:1504–1510.
48. Timcke R, von Leden H, Moore P. Laryngeal vibrations: measurements of the glottic wave: part I. The normal vibratory cycle. *AMA Arch Otolaryngol*. 1958;68:1–19.
49. Holmberg EB, Hillman RE, Perkell JS. Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice. *J Acoust Soc Am*. 1988;84:511–529.
50. Schlegel P, Stingl M, Kunduk M, et al. Dependencies and Ill-designed parameters within high-speed videoendoscopy and acoustic signal analysis. *J Voice*. 2019;33:811.e811–811.e812.
51. Collyer S, Davis PJ, Thorpe CW, et al. Sound pressure level and spectral balance linearity and symmetry in the messa di voce of female classical singers. *J Acoust Soc Am*. 2007;121:1728–1736.
52. Kreiman J, Gerratt BR, Antoñanzas-Barroso N. Measures of the glottal source spectrum. *J Speech Lang Hear Res*. 2007;50:595–610.
53. Sundberg J. Objective characterization of phonation type using amplitude of flow glottogram pulse and of voice source fundamental. *J Voice*. 2022;36:4–14.
54. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977:159–174.
55. Herbst CT, Schutte HK, Bowling DL, et al. Comparing chalk with cheese-the EGG contact quotient is only a limited surrogate of the closed quotient. *J Voice*. 2017;31:401–409.
56. Björklund S, Sundberg J. Relationship between subglottal pressure and sound pressure level in untrained voices. *J Voice*. 2016;30:15–20.
57. Sundberg J, Andersson M, Hultqvist C. Effects of subglottal pressure variation on professional baritone singers' voice sources. *J Acoust Soc Am*. 1999;105:1965–1971.
58. Sundberg J, Fahlstedt E, Morell A. Effects on the glottal voice source of vocal loudness variation in untrained female and male voices. *J Acoust Soc Am*. 2005;117:879–885.
59. Kumar SP, Švec JG. Kinematic model for simulating mucosal wave phenomena on vocal folds. *Biomed Signal Process Control*. 2019;49:328–337.
60. Patel RR, Sundberg J, Gill B, et al. Glottal airflow and glottal area waveform characteristics of flow phonation in untrained vocally healthy adults. *J Voice*. 2022;36. 140.e141-140.e121.
61. Herbst CT, Ternström S, Švec JG. Investigation of four distinct glottal configurations in classical singing—a pilot study. *J Acoust Soc Am*. 2009;125:EL104–EL109.
62. Lamesch S, Doval B, Castellengo M. Experimental study of the frequency leap interval produced by the change of laryngeal vibratory mechanism during sustained notes. In: Bresin R, Askenfelt A, eds. *Proceedings of SMAC 2013, 4th Stockholm Music Acoustics Conference*. Stockholm, Sweden: KTH Royal Institute of Technology; 2013:280–285.
63. Miller DG, Švec JG, Schutte HK. Measurement of characteristic leap interval between chest and falsetto registers. *J Voice*. 2002;16:8–19.

Supplement D: Paper IV

Simulated Laryngeal High-Speed Videos for the Study of Normal and Dysphonic Vocal Fold Vibration.

P. AICHINGER, S. P. KUMAR, H. LEHOUX, J. G. ŠVEC

Journal of Speech, Language, and Hearing Research, **65**(7), 2431-2445, (2022)

doi: 10.1044/2022_jslhr-21-00673

ASHA American Speech-Language-Hearing Association

JSLHR

Research Article

Simulated Laryngeal High-Speed Videos for the Study of Normal and Dysphonic Vocal Fold Vibration

Philipp Aichinger,[a] S. Pravin Kumar,[b] Hugo Lehoux,[c] and Jan G. Švec[c,d]

[a] Division of Phoniatrics-Logopedics, Department of Otorhinolaryngology, Medical University of Vienna, Austria [b] Department of Biomedical Engineering, Sri Sivasubramaniya Nadar College of Engineering, Chennai, India [c] Voice Research Laboratory, Department of Experimental Physics, Faculty of Science, Palacký University, Olomouc, Czech Republic [d] Voice and Hearing Centre Prague, Medical Healthcom, Ltd., Czech Republic

ARTICLE INFO

Article History:
Received December 22, 2021
Revision received February 22, 2022
Accepted March 29, 2022

Editor-in-Chief: Bharath Chandrasekaran
Editor: Jack J. Jiang

https://doi.org/10.1044/2022_JSLHR-21-00673

ABSTRACT

Purpose: Laryngeal high-speed videoendoscopy (LHSV) has been recognized as a highly valuable modality for the scientific investigations of vocal fold (VF) vibrations. In contrast to stroboscopic imaging, LHSV enables visualizing aperiodic VF vibrations. However, the technique is less well established in the clinical care of disordered voices, partly because the properties of aperiodic vibration patterns are not yet described comprehensively. To address this, a computer model for simulation of VF vibration patterns observed in a variety of different phonation types is proposed.
Method: A previously published kinematic model of mucosal wave phenomena is generalized to be capable of left–right asymmetry and to simulate endoscopic videos instead of only kymograms of VF vibrations at single sagittal positions. The most influential control parameters are the glottal halfwidths, the oscillation frequencies, the amplitudes, and the phase delays.
Results: The presented videos demonstrate zipper-like vibration, pressed voice, voice onset, constant and time-varying left–right and anterior–posterior phase differences, as well as left–right frequency differences of the VF vibration. Video frames, videokymograms, phonovibrograms, glottal area waveforms, and waveforms of VF contact area relating to electroglottograms are shown, as well as selected kinematic parameters.
Conclusion: The presented videos demonstrate the ability to produce vibration patterns that are similar to those typically seen in endoscopic videos obtained from vocally healthy and dysphonic speakers.
Supplemental Material: https://doi.org/10.23641/asha.20151833

Vocal fold (VF) geometry and kinematics reflect vocal health status, and they serve as key elements in relating voice physiology to voice acoustics and perception. In contrast to self-oscillating models, kinematic VF models avoid the need of solving tissue–flow–acoustic interactions but still support the understanding of VF vibratory behavior by implementing empirically based rules of tissue oscillation. Moreover, kinematic models combined with their visualization through images and videos enable investigation of the diverse effects of varying kinematic parameters on VF vibration. Knowledge obtained from such investigation is useful for gaining intuition of these effects. In addition, such models provide means for assisting the interpretation of clinical laryngoscopic videos.

In an early kinematic model of VF vibration, two planes that were mirror symmetric over the median plane were used as the VFs' medial surfaces and driven sinusoidally (Titze, 1984). The planes were later curved to better fit anatomy (Cranen & Schroeter, 1996; Titze & Alipour, 2006). The model was combined with a vocal tract model to produce audio signals (Story, 2013) and generalized to left–right asymmetry found in unilateral VF paralysis (Samlan & Story, 2017). Three-dimensional (3D) geometry of VFs was modeled using parametric curved polynomial surfaces based on magnetic resonance imaging of human larynges (Wu & Zhang, 2016).

Correspondence to Philipp Aichinger: philipp.aichinger@meduniwien.ac.at. ***Disclosure:*** *The authors have declared that no competing financial or nonfinancial interests existed at the time of publication.*

A salient milestone in kinematic modeling of VF vibration was the development of the phase-delayed overlapping sinusoid (PDOS) model (Titze, 1988; Titze & Alipour, 2006), via revisiting the earlier proposed sinusoidal driving of VF trajectories (Titze, 1984). PDOS uses phase-delayed sinusoids to model trajectories of the upper and lower margins of the VFs. Compared with the two-mass model, the phase delay is controllable more easily in the kinematic PDOS model. The PDOS model was extended to use sums of sinusoids instead of single sinusoids, enabling nonsinusoidal cycle shapes (Jiang et al., 2000). However, in these kinematic models from the past, emphasis was neither given to generating realistically looking videos emulating the laryngoscopic view, nor to combining inferior, medial, and superior VF surfaces, nor to perturbation, nor to the vertical component of vibration. These aspects are addressed in this article.

The recently proposed kinematic model for synthesizing single-line videokymograms can be seen as another step in the development of the PDOS model (Kumar & Švec, 2019). It used sinusoids for a number of contour points of the VFs' frontal cross section. Sinusoids were combined for lateral and vertical movement, which enabled circular or elliptical motion in the frontal plane and simulation of the mucosal waves that travel upward on the medial VF surfaces and continue laterally on the top VF surfaces. These waves are physiologically important, because they make the energy transfer from the airflow to the VFs efficient and serve as clinical indicators of the health of the VF mucosa. The kymogram synthesizer had been later generalized to be capable of simulating time-constant left-to-right phase differences, and it was validated by fitting simulated kymograms to clinical kymograms (Bulusu et al., 2021).

The purpose of this study is to expand the 2D model of Kumar and Švec (2019) from a single coronal section to a 3D kinematic model of the entire VFs. Combined with the simulated laryngoscopic visualization, this expansion allows generating high-speed videos in addition to the kymograms and is capable of simulating laryngeal high-speed videos (LHSVs) of a variety of phonation types, which are known from experimental and clinical studies. The first aim is to implement conditions that vary in terms of glottal adduction, from zipper-like vibration found in breathy voice to pressed phonation with an excessively short open phase. The transition between the two extremes enables generating conditions for breathy voice onset combined with a pressed sustained phase of phonation. The second aim is to implement glottal conditions with left–right and anterior–posterior differences that do not vary over time. In particular, time-constant phase and frequency differences across the lateral and sagittal[1] dimension were explored. The third aim is to implement conditions for highly irregular phonation showing phase differences that vary on the pulse-to-pulse time scale.

The remainder of this article is structured as follows. In the Method section, the computer model for simulating LHSVs is described. In the Results section, a selection of nine videos is described to demonstrate the capabilities of the synthesizer. In the Discussion and Conclusions sections, added values of our approach with regard to the prior state of the art are listed, and limitations and suggestions for future work are provided.

Method

We synthesize the LHSVs by concatenating a number of simulated kymograms obtained with an updated version of the model proposed earlier (Kumar & Švec, 2019). In particular, kymograms sampled at 256 equidistant sagittal positions are used. The model is advanced here to be capable of more general left–right differences. Regarding anterior–posterior differences, a few rules with respect to the variation of kinematic VF parameters across sagittal positions are proposed. Most notably, vibration is controlled at a few spatially separate supporting points (left, right, anterior, midsagittal, and posterior) while ensuring spatial continuity of the parameters along the sagittal axis by interpolation. As a result, tangible control parameters are provided that enable developing an intuition of how they relate to VF vibration patterns. Anterior–posterior differences of kinematic parameters are introduced such that the model is capable of producing a few qualitative key features that are observed in the Vienna database of LHSV (Aichinger et al., 2016) and described elsewhere (Aichinger & Pernkopf, 2021; Childers et al., 1986; Orlikoff et al., 2012; Shiba & Chhetri, 2016; Verdonck-de Leeuw et al., 2001; Yokonishi et al., 2016).

For identifying the features observed in the LHSV database and setting the parameters of the kinematic model, the first author P.A. (with 10 years of experience with LHSV) thoroughly reviewed the clinical videos using various modalities of visualization, such as single- and multiline digital videokymography, phonovibrograms, glottal area waveforms, and frame montages. He then manually adjusted the model parameters so that the synthetic videos demonstrated the specific clinical features (i.e., the zipper-like vibrations, pressed phonation, breathy voice onset, left–right phase differences, anterior–posterior phase differences, left–right frequency differences, and left–right and anterior–posterior jitter differences). Because these clinical features are purposefully directly linked to the adjustable model parameters (i.e., glottal halfwidth, initial vibratory phases of the left and right VFs, anterior–posterior phase delays, vibratory frequencies of the left and right VFs, left

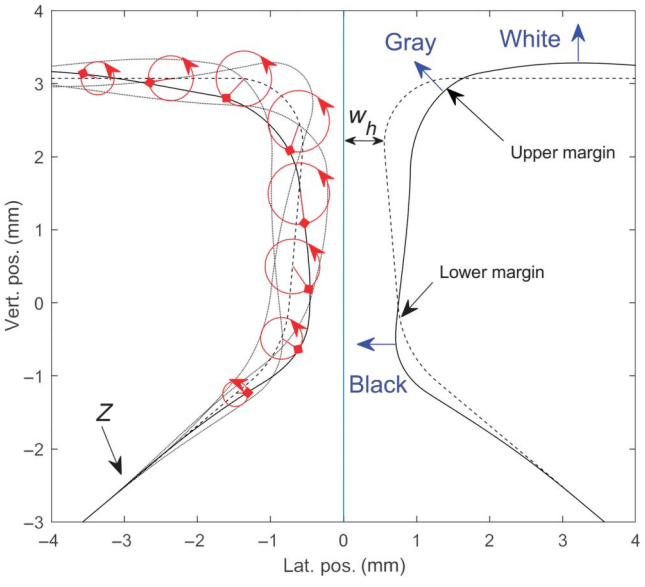
[1]The anterior–posterior direction has often been referred to as longitudinal, which normally refers to the anatomical orientation of the spine. Thus, we prefer the term *sagittal* over *longitudinal*.

and right jitter, and anterior–posterior jitter differences), it proved relatively straightforward to set the model parameters to qualitatively demonstrate the targeted clinical features. The authors J.G.S., H.L., and S.P.K. (with over 20, 3, and 5 years of experience with LHSV, respectively) reviewed the synthesized videos and confirmed their resemblance to clinical videos.

Generation of Single-Line Kymograms

The generation of single-line kymograms is based on a kinematic model proposed earlier (Kumar & Švec, 2019). Figure 1 illustrates the reference geometry of the VFs, VF vibration, and the model of local illumination. The reference geometry of the VFs represented by the contours in a frontal plane is initialized according to the M5 model (Scherer et al., 2001). The M5 parameters varied in this study are the glottal halfwidth w_h, which reflects the opening of the glottal reference geometry; the convergence angle α, which reflects the declination of the medial surface in the reference geometry; and the vertical thickness of the VFs. The reference geometry is the VFs' contours at zero amplitude of vibration representing the prephonatory shape also referred to as the "resting position." Surface points of the VFs are sampled from the M5 contours with a resolution of 25 μm. To simulate vibration, the surface points are moved circularly. Circular trajectories are used here for simplicity, although elliptical motions are also possible in the model. The vibration amplitudes are the circles' radii, which are imposed on the lower and the upper margin separately. At the bottom VF surface, the amplitudes of the points below the upheaval point Z are set to zero. Amplitudes in between these three supporting points, that is, Z, the lower margin, and the upper margin, are interpolated. Phase differences between individual surface points are imposed to simulate the mucosal waves that are surface waves traveling from the subglottal upheaval point across the VF surface upward and laterally. At the top surface, the amplitude decays gradually toward the lateral end, resulting in damping of the outward travelling mucosal waves, owing to the mucosal wave extent parameter as described earlier (Kumar & Švec, 2019). In particular, the mucosal wave extent reflects how far the mucosal wave travels beyond the glottal edge on the superior surface of the VFs. Smaller mucosal wave extent is simulated by faster decay of the amplitude of the mucosal wave travelling across the VF surface laterally.

Figure 1. Illustration of the contours of the vocal folds in a frontal section, including left–right differences. The reference geometry (dashed curves) is obtained using the M5 model (Scherer et al., 2001). Contours at an example time instant (solid curves) differ between the left and the right vocal folds. Dotted curves on the left are contours at equidistant points in time. Red circles indicate trajectories of example contour points. Their midpoints coincide with the M5 reference contour. Red diamond-headed lines indicate current vibratory phases. Note that the current phases increase clockwise from bottom to top. Their evolvement over time (red arrows) is counterclockwise, resulting in mucosal waves traveling upward and outward. Vert. = vertical; pos. = position; Lat. = lateral.



Furthermore, the Kumar & Švec model is generalized here to enable a variety of left–right differences including cycle frequency perturbations commonly referred to as frequency jitter. Shimmer is applied by modulating amplitudes randomly from one pulse to the next.

Kymograms are obtained from the vibrating VF contours described above using a local illumination model based on diffuse reflection as proposed in Kumar and Švec (2019). In particular, light intensity across the VF surface is proportional to the distance to the light source, as well as the slope of the contour. The slope is the surface declination with regard to the direction of light incidence.

The kinematic VF parameters that are allowed to be different for the left and the right VF are listed as follows.

1. *Left–right differences in frequency*
   To simulate diplophonia, sometimes also referred to as biphonation, different vibration frequencies for the left and the right VF are used. Typically, due to coupling via the glottal airstream as well as collision, the frequencies are small integer multiples of a common cycle frequency, for example, 3/4 or 4/5 (Lucero et al., 2015).
2. *Left–right differences in phase*
   Constant and time-varying phase differences are distinguished. To simulate a time-constant delay of one VF with regard to the other, a time-constant left–right phase difference is imposed as proposed in the work of Bulusu et al. (2021). To simulate irregular VF vibration, a time-variant left–right phase difference is imposed. The phase difference is allowed to

vary from one pulse to the next. In particular, the difference is imposed by randomly shifting times of individual pulses, as proposed in the work of Aichinger and Pernkopf (2021). Time shifts are Gaussian random numbers and differ between the left and the right VFs. This results in a jitter that is different for the two VFs.

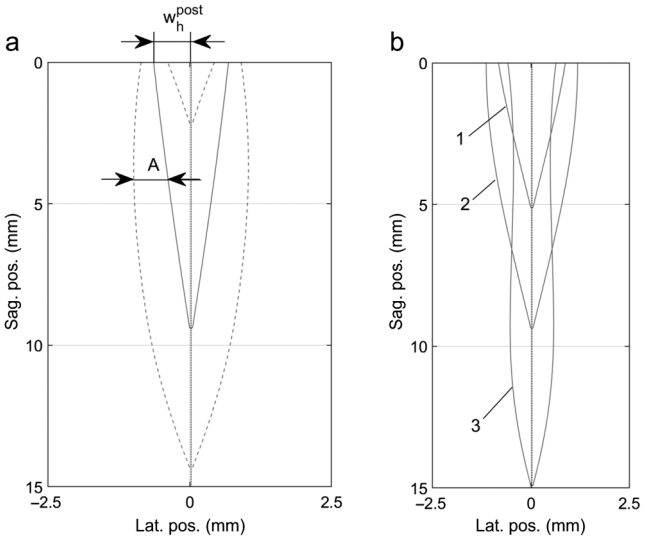
3. *Left–right differences in amplitude*
   To simulate left–right amplitude asymmetry, the vibration amplitudes A are enabled to differ between the left and the right VF.

Generation of Simulated Laryngoscopic Videos by Combining Single-Line Kymograms

For each simulated LHSV, 256 kymograms at equidistant sagittal positions are generated and concatenated. The first kymogram is located at the anterior end of the VFs, which is the anterior commissure, and the last one at the posterior end corresponding to the vocal processes. The length of the VFs is 15 mm in all presented simulations. Variation of kinematic parameters across sagittal positions is described as follows.

1. *Anterior–posterior difference in glottal halfwidth*
   The posterior glottal halfwidth w_h^{post} illustrated in Figure 2a controls the opening of the reference glottal contours at the posterior end of the VFs. At the anterior end, the glottal halfwidth is negative to compensate for the upper margin radius of the M5 model, which makes the top surface of the VFs plain where the VFs are adnate. At sagittal positions in between the anterior and the posterior ends, the glottal halfwidth is linearly interpolated, making the reference VFs' margins straight lines (V-shaped solid contours as shown from above in Figure 2a).
2. *Anterior–posterior difference in vibration amplitudes*
   Vibration amplitudes A are controlled at the midsagittal position and at the posterior end separately, whereas they are zero at the anterior end. They are obtained by shape-preserving piecewise cubic interpolation in between, which ensures continuity. The vibration amplitude is assumed to be maximal at the midsagittal position and smaller at the posterior end. For example, when the amplitude is zero at the posterior end, the shape of the amplitude is approximately half-sinusoidal (see, e.g., Figure 4d). Figure 2a illustrates the vibration amplitude with regard to the V-shaped margins at resting position (solid lines) and excursion extremes (dashed lines).
3. *Anterior–posterior differences in phase*
   Figure 2b illustrates waves that travel on the VF margins in a sagittal direction. This is achieved by imposing an anterior–posterior phase difference Φ_s. The phase difference is controlled at the anterior end with respect to the posterior end. At sagittal positions in between, the phase difference is obtained by linear interpolation.

Figure 2. Illustrations of the vocal fold margins as shown from above. (a) The posterior glottal halfwidth w_h^{post} is indicated with regard to the V-shaped glottal margins (solid lines) at resting position. Dashed lines show glottal margins at excursion extremes defined by the vibration amplitude A. (b) Illustration of sagittal phase differences. The numbers 1, 2, and 3 refer to margins at three subsequent time instances. Mirror-symmetrical waves travel from posterior (top) to anterior (bottom). Lat. = lateral; pos. = position; Sag. = sagittal.



Similar to what is done for left–right phase differences, we distinguish constant and time-varying sagittal phase differences. To simulate irregular VF vibration, a time-varying anterior–posterior phase difference Φ_s is imposed. The phase difference is allowed to vary from one glottal pulse to the next, resulting in a jitter that is different at all sagittal positions. In particular, the phase difference is obtained by randomly modulating pulse times at the anterior and the posterior end separately and independently of each other, as described above for left–right phase differences. The phase is interpolated linearly across sagittal positions between the ends. The randomly time-varying Φ_s is either an identical or a different realization of the random process for the left and the right VF, resulting in left–right symmetric or asymmetric anterior–posterior phase differences, respectively.

Results

Table 1 gives an overview of the synthesis parameters used to generate the individual videos. The values of

Table 1. Synthesis parameters used to generate the presented Supplemental Videos V1–V9.

Parameters	Videos								
	V1	V2	V3	V4	V5	V6	V7	V8	V9
f_o	230	70	230	115	230	Left: 180, right: 240	230	230	230
α	10	10	20	10	10	10	10	10	10
T	3	3	3	3	2	3	2	2	2
Φ_v	90	60	60	20	10	20	10	10	10
Φ_l	0	0	0	90	0	0	Rand.var.	0	Rand.var.
Φ_s	0	0	0	0	120	0	0	Rand.var.	Rand.var.
E	50	5	3	5	3	5	3	3	3
w_h^{post}	1	0	0	1	0.1	0.5	1	.7	.7
$A_{u,l}^{post}$	0.1, 0.1	0	0.6, 0.5	0	0	Left: 0.4, 0.3 Right: 0.6, 0.4	0.2, 0.1	0.1, 0.05	.1,.05
$A_{u,l}^{mid}$	0.3, 0.2	1, 0.8	0.7, 0.6	0.6, 0.4	0.7, 0.5	Left: 0.3, 0.2 Right: 0.5, 0.3	0.6, 0.4	0.3, 0.2	.3,.2

Note. Fundamental frequency f_o (Hz), convergence angle α (°), vertical thickness T (mm), vertical phase difference Φ_v (°), lateral phase difference Φ_l (°), sagittal phase difference Φ_s (°), mucosal wave extent E, posterior glottal halfwidth w_h^{post} (mm), and the posterior and midsagittal vibration amplitudes of the upper and lower margins $A_{u,l}^{post}$ and $A_{u,l}^{mid}$ (mm). V1 = zipper-like vibration; V2 = pressed phonation; V3 = voice onset; V4= time-constant left–right phase differences; V5 = time-constant anterior–posterior phase differences; V6 = left–right diplophonia; V7 = time-varying left–right phase differences; V8 = time-varying anterior–posterior phase differences; V9 = time-varying left–right and anterior–posterior phase differences; Rand.var. = randomly varying parameters.

the parameters were determined experimentally such that the simulated videos look similar to clinical ones contained in the Vienna database of pathological and nonpathological voices (Aichinger et al., 2016).

Zipper-Like Vibrations

Supplemental Video V1 shows a zipper-like vibration pattern. Figure 3 shows a video frame at maximal opening (a) and a frame at minimal opening (b). The VF margins are concave and convex, respectively, as shown from above. The zipper-like vibration is characterized by the movement of the most posterior collision point C. C moves anteriorly during the opening phase and posteriorly during the closing phase. This may be reminiscent of the zipper of a jacket moved down and up in a cyclic way, which justifies the use of the term *zipper-like* proposed earlier (Childers et al., 1986; Orlikoff et al., 2012). A posterior glottal chink is observed, due to the posterior glottal halfwidth ($w_h^{post} = 1$ mm) that is larger than the vibration amplitude at the posterior end (top of the frames, A^{post} = 0.1 mm). The zipper-like movement occurs due to the triangular prephonatory shape of the glottis, as zero anterior–posterior phase differences were prescribed for this case. Figure 3c shows the glottal halfwidth with respect to sagittal position, and the upper and lower VF margin vibration amplitudes. At the anterior commissure (bottom of the figure), w_h is negative to compensate the rounding of the upper margin contour. The glottal halfwidth increases linearly in the posterior direction until it reaches its maximum w_h^{post} at the posterior end (top of the figure). The vibration amplitudes are zero at the anterior end. They equal their nominal values A^{mid} and A^{post} at midsagittal position and at the posterior end, respectively, and are obtained by shape-preserving piecewise cubic interpolation in between.

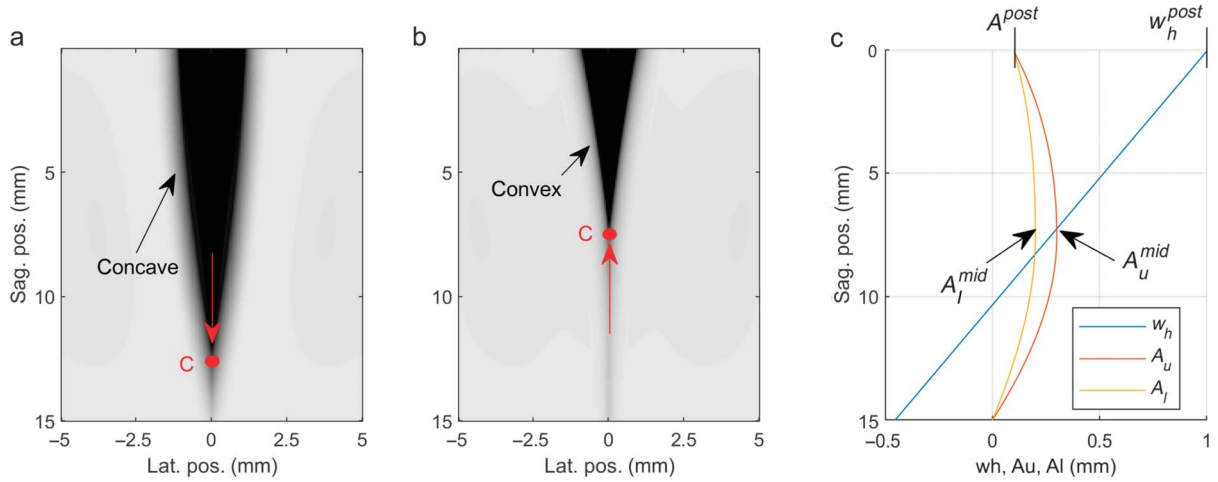
Pressed Phonation

Supplemental Video V2 shows VF vibration of pressed phonation similar to those described by Yokonishi et al. (2016). Figure 4 shows a video frame at maximal opening (a), a frame at maximal closure (c), and a frame in between (b). In (a), the peaks of the mucosal waves are laterally close to the VF margins. The peaks are located at the midsagittal position, because the vibration amplitudes are maximal there, see (d). In the middle frame (b), the mucosal waves have travelled laterally. Their amplitudes decay over time toward lateral sides. In frame (c), they have vanished completely. Pressed phonation is characterized by a negative glottal halfwidth and large amplitudes, see (d). The open quotient in the shown example is approximately 30%.

Breathy Voice Onset

Supplemental Video V3 shows a breathy voice onset transitioning into pressed phonation (Shiba & Chhetri, 2016). Kinematic parameters over time such as the additional posterior glottal halfwidth w_h^{post+} and the amplitude factor A^* for oscillation fade-in are shown in Figure 5a. w_h^{post+} is added to w_h^{post}, which is the residuum posterior glottal halfwidth during sustained phonation. All vibration amplitudes are multiplied by A^* that goes from 0 to 1

Figure 3. Illustration of zipper-like vocal fold vibration (Supplemental Video V1). At a video frame of maximal opening (a), the vocal fold margins are concave, and at minimal opening (b), they are convex, as shown from above. The most posterior collision point C moves up and down like a zipper. (c) Selected kinematic parameters with respect to the sagittal position (0 mm: posterior end, 15 mm: anterior end). Glottal halfwidth w_h, vibration amplitudes of the upper and lower VF margins A_u and A_l, respectively. Sag. = sagittal; pos. = position; lat. = lateral.



over time; thus, the amplitudes start at 0 and approach their nominal values.

The resulting kymographic vibratory pattern obtained at the midsagittal position is shown in Figure 5b. In the beginning, the VFs are open and do not vibrate. There, the additional posterior halfwidth w_h^{post+} is equal to 2 mm, and the amplitude factor A^* is equal to 0. Adduction and vibration start at 4 ms approximately. Parameters w_h^{post+} and A^* transit smoothly to 0 and 1, respectively, after which phonation is sustained. The phonovibrogram in Figure 5c is a false color illustration of the deflections of the VF margins with respect to sagittal position and time (Lohscheller et al., 2008; Lohscheller & Eysholdt, 2008). The sagittal position in the middle of the image corresponds to the posterior end, and the top and bottom correspond to the anterior end. The upper half of the image corresponds to the left VF, and the lower half is flipped upside down and corresponds to the right VF, as suggested in the works of Lohscheller et al., (2008) and Lohscheller and Eysholdt (2008). Deflections (green) refer to distances of the VF margins from the midline that goes from anterior to posterior. Figure 5d shows the resulting

Figure 4. Illustration of pressed vocal fold vibrations (Supplemental Video V2). Shown are frames of maximal opening (a), maximal closure (c), and a frame in between (b). The mucosal waves traveling laterally. (d) Selected kinematic parameters with respect to sagittal position, glottal halfwidth w_h, vibration amplitudes of the upper and lower VF margins A_u and A_l. Sag. = sagittal; pos. = position; Lat. = lateral.

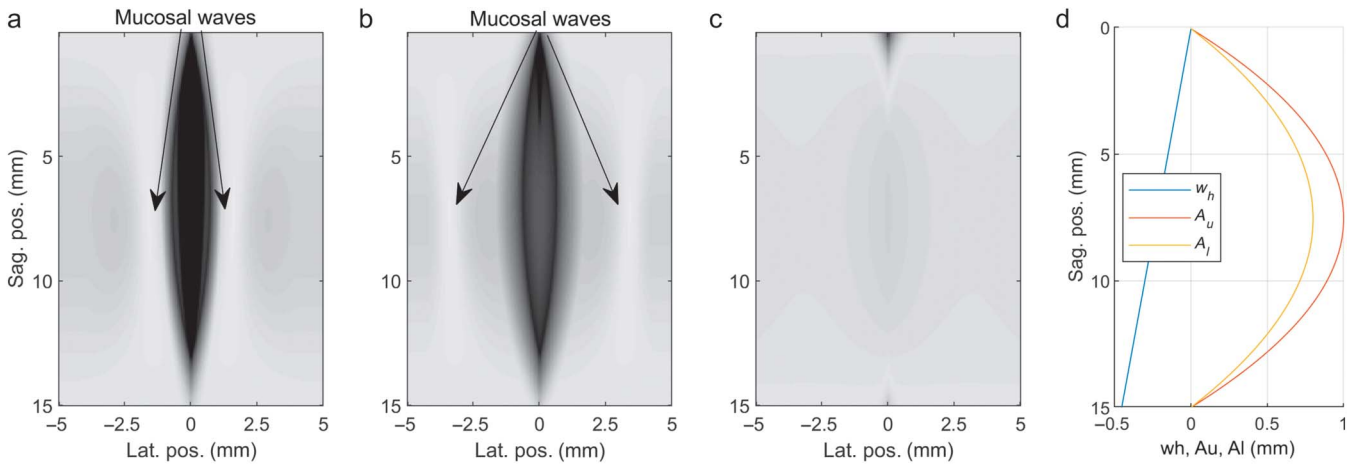
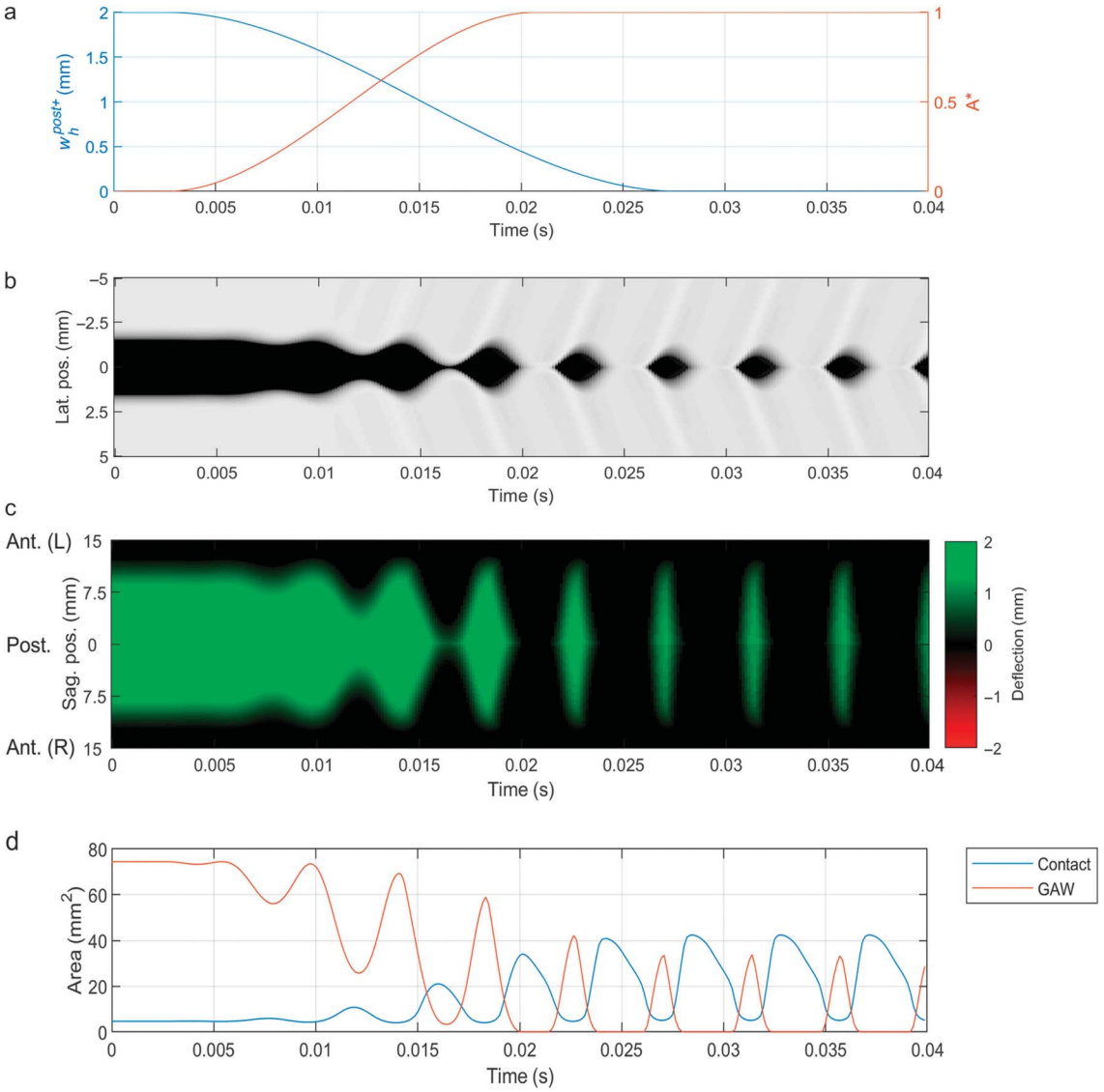


Figure 5. Illustration of a voice onset (Supplemental Video V3). (a) Kinematic control parameters over time. Shown are the additional posterior opening w_h^{post+} and the amplitude modulator A^*. (b) Kymogram. (c) Phonovibrogram. (d) Contact area reflecting the electroglottogram (EGG) and glottal area waveform (GAW). Lat. = lateral; pos. = position; Ant. = anterior; Post. = posterior; Sag. = sagittal; L = left; R = right.



VF contact area waveform (electroglottogram [EGG] signal) and the GAW.

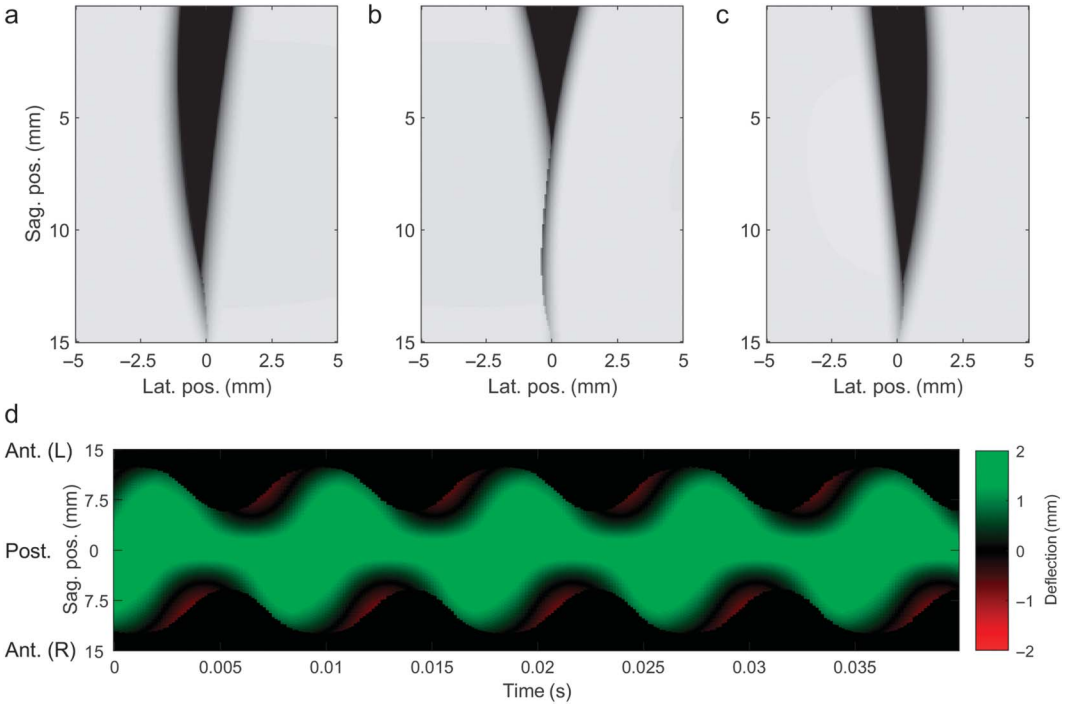
Left–Right Phase Difference

Supplemental Video V4 shows VF vibrations involving a time-constant left–right phase difference (Verdonck-de Leeuw et al., 2001). Figure 6 shows frames in which the deflection of the right and left VFs is maximal, see (a) and (c), respectively, and a frame halfway between (b). The shape of the glottal gap is V-like in (b) and strained to the left and right in (a) and (c). Figure 6d shows the phonovibrogram. Negative deflections due to trespassing of the margins toward the opposite side are shown in red. The phase difference is 90°.

Anterior–Posterior Phase Difference

Supplemental Video V5 shows VF vibrations involving a time-constant anterior–posterior phase delay (Orlikoff et al., 2012). Figure 7 shows video frames just after the start of the opening (a), at maximal opening (b), and just before closure (c). The phase delay with respect to sagittal position is shown in (d). Waves traveling from

Figure 6. Illustration of vocal fold vibrations involving a time-constant left–right phase difference of 90° (Supplemental Video V4). Shown are frames of maximal deflection of (a) the right and (c) the left vocal folds, and (b) a frame halfway in between. (d) Phonovibrogram. Lat. = lateral; pos. = position; Ant. = anterior; Post. = posterior; Sag. = sagittal; L = left; R = right.



posterior to anterior are observed. Figure 7e shows multiline kymograms for posterior, midsagittal, and anterior positions (top to bottom). Red vertical lines indicate times of maximal midsagittal opening. Maximal posterior openings occur earlier (Ea), and maximal anterior openings later (La). The anterior–posterior phase delay is reflected in the phonovibrogram by tilts of the green areas to the right; see Figure 7f. The phase difference is 120°.

Left–Right Frequency Difference

Supplemental Video V6 shows VF vibrations of left–right diplophonia often referred to as biphonation, in which the left and the right VFs vibrate at different frequencies (Aichinger & Pernkopf, 2021). Figure 8a shows the digital videokymogram, where black vertical lines indicate cycle boundaries that coincide with the time instants of minimal glottal width. In each cycle, the left VF has three excursions, whereas the right VF has four. This reflects the left–right frequency ratio of 3/4. Figure 8b shows the phonovibrogram, in which white vertical lines indicate cycle boundaries. A 3/4 pattern and negative deflections crossing the glottal midline (red) are observed. Figure 8c shows the contact area waveform reflecting the EGG, and the GAW, which have an approximate antiphase relationship. Three peaks are observed within each cycle in the contact area waveform, but four peaks are observed in the GAW.

Left–Right Jitter Differences

Supplemental Video V7 shows VF vibration involving jitter that is different for the left and the right VF. In particular, left–right phase differences vary randomly over time, from one pulse to the next. Figure 9a shows the phase distortion of the left and the right VF, that is, Φ^{L}_{dist}, and Φ^{R}_{dist}. They are added to the vibratory phases. Phase distortions larger and smaller than zero correspond to temporal anticipation and delay of VF vibration, respectively. In particular, when the phase distortion is larger on the right than on the left, the left VF is delayed with respect to the right one and vice versa (see phonovibrogram in Figure 9b). Switches of the orientation of the delay are indicated by vertical blue lines in Figures 9a and 9b. Figure 9c shows the contact area waveform and the GAW. Shimmer, which is a random amplitude modulation from one pulse to the next, is also applied. Random peak height fluctuations observed in the contact area waveform and the GAW result from random

Figure 7. Illustration of a simulated anterior–posterior phase difference of 120° (Supplemental Video V5). Video frames just after the start of the opening (a), at maximal opening (b), and just before closure (c). (d) Anterior–posterior phase delay. (e) Multiline kymograms, posterior (top), midsagittal (middle), anterior (bottom). Red lines reflect times of maximal midsagittal opening. Times of maximal posterior and anterior widths are labeled as early (Ea) or late (La), with respect to red lines. (f) Phonovibrogram. Lat. = lateral; pos. = position; Ant. = anterior; Post. = posterior; Sag. = sagittal; L = left; R = right.

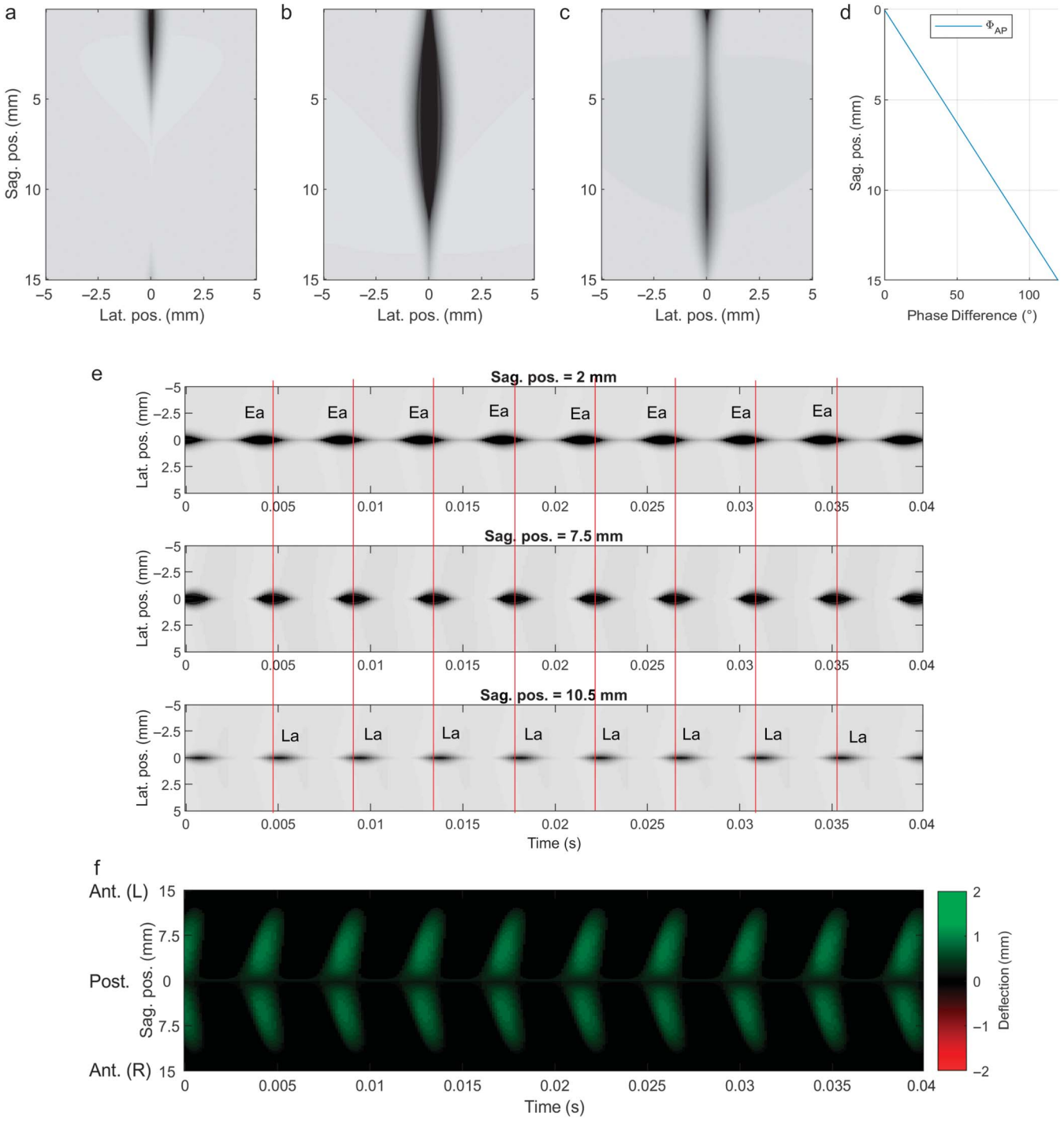
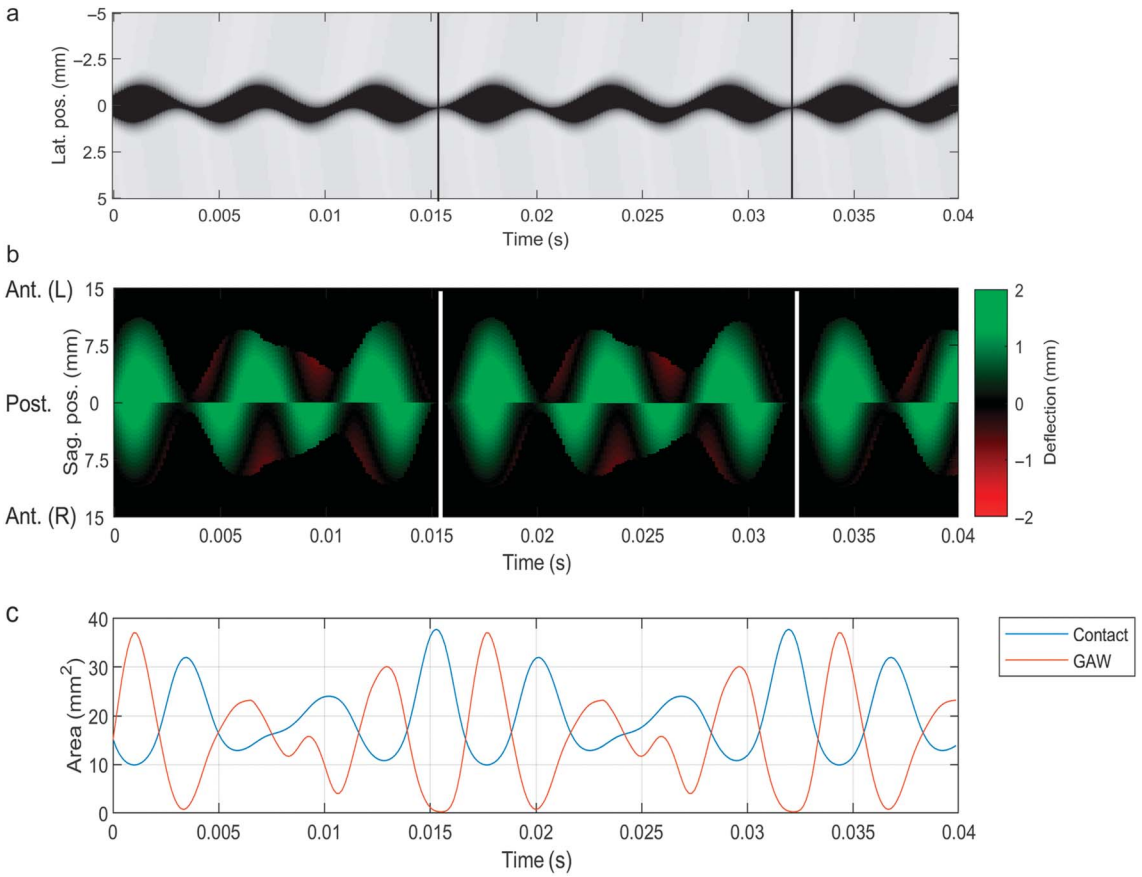


Figure 8. Illustration of left–right diplophonia, in which the vocal folds vibrate at different frequencies (Supplemental Video V6). (a) Videokymogram. The frequency ratio of the left and the right vocal fold is 3/4. Black vertical lines indicate cycle boundaries coinciding with times of minimal glottal width. (b) Phonovibrogram. White vertical lines indicate cycle boundaries. (c) Contact area waveform reflecting the electroglottogram and glottal area waveform (GAW). Lat. = lateral; pos. = position; Ant. = anterior; Post. = posterior; Sag. = sagittal; L = left; R = right.



modulation of amplitude, as well as from random phase differences.

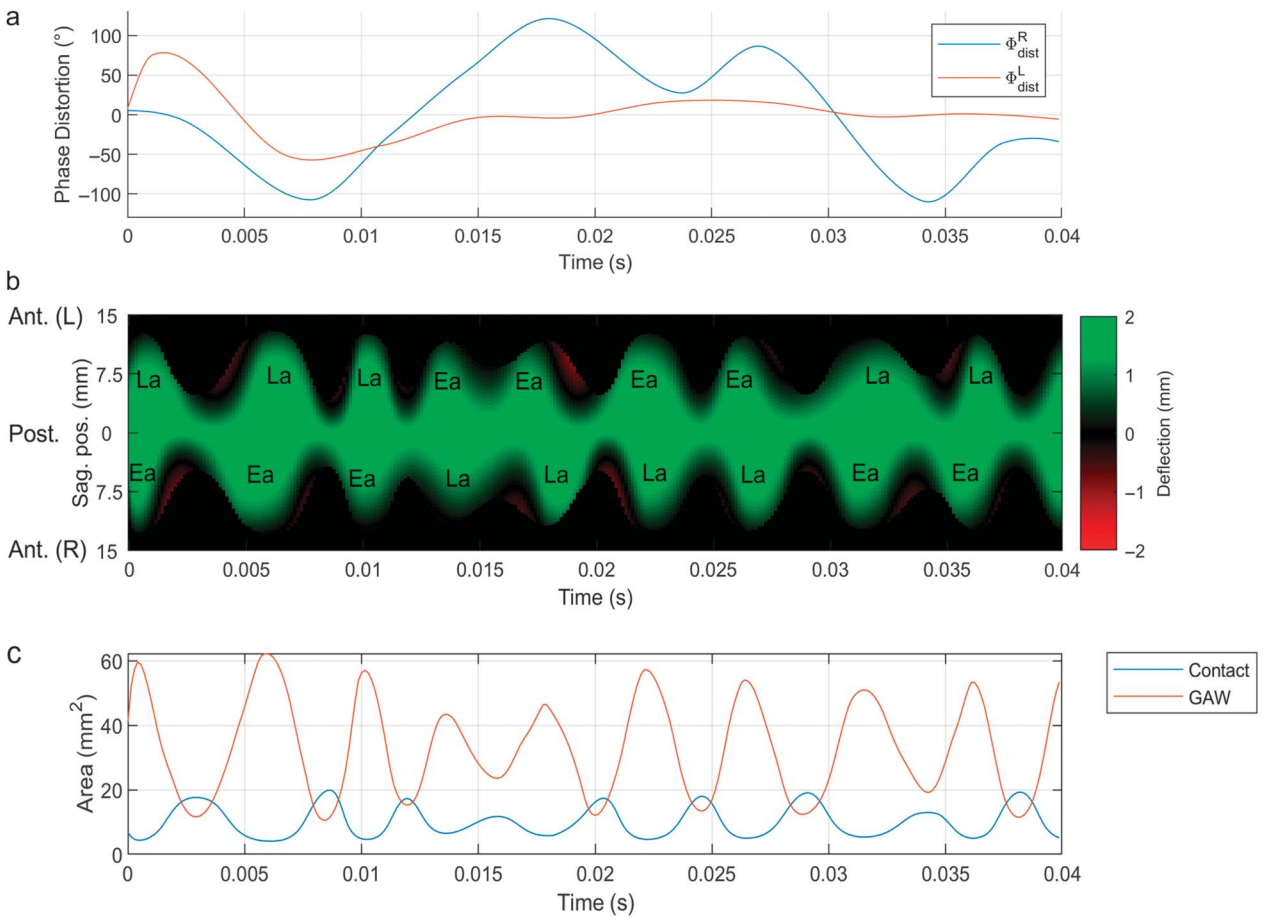
Anterior–Posterior Jitter Differences

Supplemental Video V8 shows a vibration pattern involving jitter that is different across sagittal positions but equal for the left and the right VFs. In particular, anterior–posterior phase differences vary randomly over time. As a result, the travelling waves observed on the VFs' margins travel anteriorly or posteriorly, according to the current sign of the time-varying phase difference. Figure 10a shows random phase distortions for the posterior and the anterior end over time. Figure 10b shows the multiline kymogram for posterior (top), midsagittal (middle), and anterior positions (bottom). Red vertical lines indicate instances of maximal midsagittal opening. Instances of maximal posterior and anterior opening are labeled as early (Ea), in time (In) or late (La) with respect to midsagittal maxima. Anterior pulsation is early or late when posterior pulsation is late or early, respectively, because the phase distortion is interpolated linearly across sagittal positions. The phonovibrogram (Figure 10c) shows pulse-to-pulse shape differences of green areas. Left–right symmetry of the VF vibrations is reflected by mirror symmetry across the horizontal line corresponding to the posterior end (Sag. pos. = 0 mm). Shimmer is also applied.

Left–Right, Anterior–Posterior Jitter Differences

Supplemental Video V9 shows a vibration pattern involving jitter that is not only different for the anterior and the posterior end but also for the left and the right VF. This is the most complex vibration pattern presented here. Figure 11 shows the phonovibrogram, in which shape differences of the green areas are observed on a pulse-to-pulse timescale. In addition, a different pattern is observed on the left (top half) and right VFs (bottom

Figure 9. Illustration of left–right phase differences that vary randomly over time (Supplemental Video V7). (a) Phase distortions of the left and the right vocal folds. (b) Phonovibrogram. Times of left and right maxima are labeled as early (Ea) or late (La), with respect to each other. (c) Contact area waveform reflecting the electroglottogram (EGG) and glottal area waveform (GAW). Lat. = lateral; pos. = position; Ant. = anterior; Post. = posterior; Sag. = sagittal; L = left; R = right.



half). In particular, the top and bottom half are not mirror symmetric across sag. pos. = 0 mm. Shimmer is also applied.

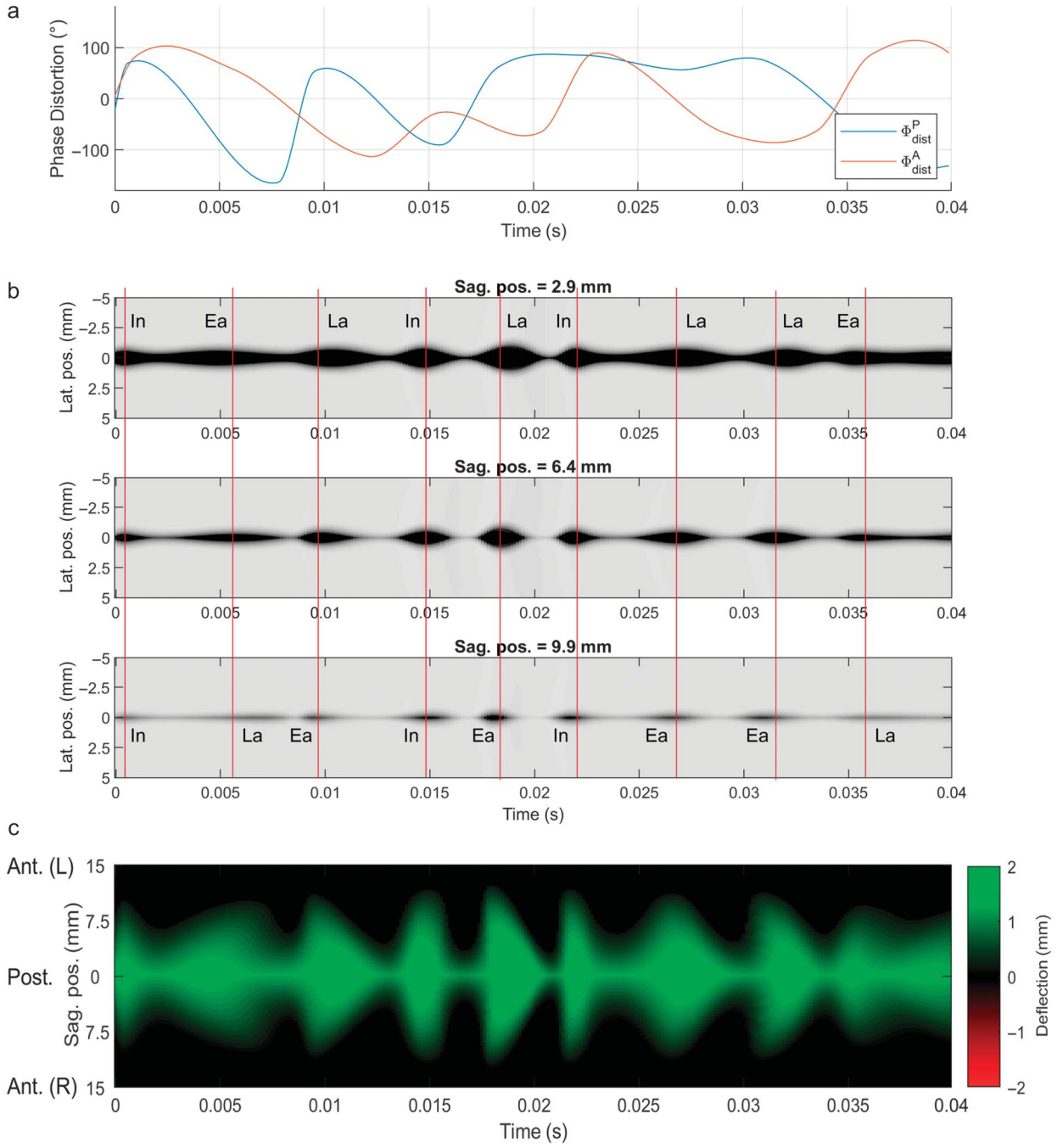
Discussion and Conclusion

A synthesizer for simulated LHSVs showing normophonic and dysphonic vibration patterns of the VFs is presented. The synthesizer is based on a formerly published model simulating videokymograms, which is extended to be capable of simulating diverse left–right and anterior–posterior differences as well as perturbations. Simulated LHSVs are obtained by concatenating simulated kymograms sampled at 256 equidistant sagittal positions. To simulate sagittal differences, kinematic VF parameters are controlled separately at a few sagittal positions and interpolated in between. Time-constant as well as time-varying parameters were used. The former enables the generation of stationary unperturbed voice quality, whereas the latter enables simulating transients, such as voice onset, and perturbations, such as jitter.

An added value of the proposed model is that it includes spatially differing perturbations. In particular, modeling of the phase distortion at different positions across the VF surfaces enables an unprecedented depth of understanding of perturbations. For example, random fluctuations of pulse heights, often referred to as shimmer, may be explained, in addition to multiplicative modulation, by randomly time-varying phase differences between spatially separate parts of the VFs as shown in Supplemental Video V7. This alternative explanation of shimmer is planned to be further investigated in the future using LHSVs.

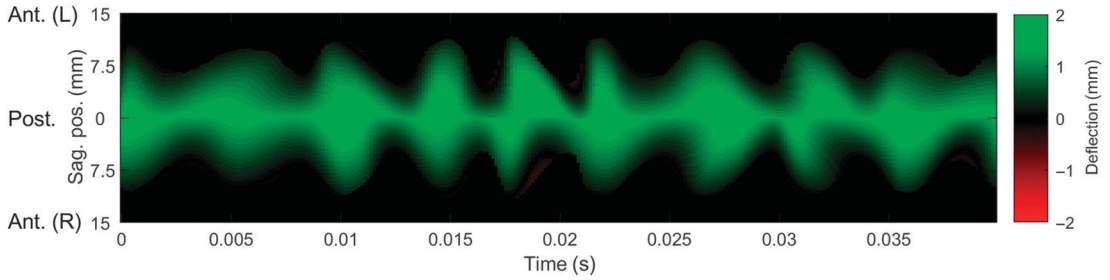
Limitations of this study and suggestions for future work include the following. The VF vibratory trajectories considered here are limited to sagittal planes due to the kymographic nature of the underlying kinematic model

Figure 10. Illustrations of anterior–posterior phase differences that vary randomly over time, but left–right mirror symmetrically (Supplemental Video V8). (a) Phase distortions at the posterior and anterior end over time. (b) Multiline kymogram for a posterior position (top), the midsagittal position (middle), and an anterior position (bottom). Red vertical lines indicate times of maximal midsagittal width. Times of maximal posterior and anterior widths are labeled as early (Ea), in time (In), or late (La), with respect to red lines. Lat. = lateral; pos. = position; Ant. = anterior; Post. = posterior; Sag. = sagittal; L = left; R = right.



(Kumar & Švec, 2019). In the future, it is desirable to implement trajectories in cross-sectional planes that are perpendicular to prephonatory positions of the individual VFs to achieve simulations that are even more realistic. The vibratory trajectories and the empirical rules for the mucosal wave speed may further be refined once more detailed experimental data from hemilaryngeal experiments (Boessenecker et al., 2007; Döllinger et al., 2016;

Figure 11. Phonovibrogram of anterior–posterior phase differences that vary randomly over time and left–right asymmetrically (Supplemental Video V9). pos. = position; Ant. = anterior; Post. = posterior; Sag. = sagittal; L = left; R = right.



Herbst et al., 2017) are acquired. For case-specific simulations, the M5-based VF shape used here could also be replaced by other individual VF shapes observed experimentally (Berry et al., 2001; Šidlof et al., 2008; Vahabzadeh-Hagh et al., 2017; Wu & Zhang, 2019).

There is considerable interest to find the kinematic model parameters that provide the best correspondence to the vibratory patterns of the VFs captured in clinical high-speed videos. However, no computational fitting of simulated videos to natural videos is provided at this point. Quantitative comparisons of the proposed simulated and clinical video data have been done on single-line kymograms so far (Bulusu et al., 2021). The same method could possibly be extended to multiline kymograms in the future. This would enable characterization of sagittal differences shown in clinical videos and provide a physiologically motivated means for parameterizing the voice source.

While the simulated vibration patterns appear to look realistic in terms of vibration of VF margins, the simulated videos are still perceptually distinguishable from clinical videos. In particular, they would not pass a Turing test, because they are idealized representations of the clinical videos. Additional visual effects that are often observed in clinical videos include differences in the intensity of illumination across the VFs, and specular reflections. The differences in illumination arise from the six degrees of freedom of positioning the light source combined with a possible shading by supraglottal structures, most notably, the epiglottis. The specular reflections of the light source show up most prominently as a manifestation of the mucosal wave travelling laterally across the VF's top surface. The specular reflections are bright patches, the shapes and movements of which appear to be hardly predictable, because small differences of the VFs' top surfaces may cause large differences in specular reflections. Besides attempting to increase the naturalness of the simulated videos via refining our physical computer model, a promising alternative for this purpose would also be to use conditional generative adversarial neural networks (GANs; Mirza & Osindero, 2014). GANs are a promising approach to generate realistically looking synthetic images by combining natural training images. GANs are data driven, whereas the approach presented in this article is model driven.

Finally, clinical videos may show (a) locally varying differences of tissue color, for example, blood vessels, bleedings, (b) mucus sticking on the VFs, (c) pathological structural varieties of the VFs, such as nodules, polyps, scars, edema, cysts, tumors, and bamboo nodes. These abnormalities were not considered here.

Possible applications of the VF vibration synthesizer include the following. First, via obtaining the GAW, the model may be used as a voice source model with audio speech synthesizers. This will enable studying the effects of physiological parameters of VF vibration on the speech audio and auditory perception. Particular features of the VF vibrations may be varied for investigating causal dependencies of the sound and auditory perceptual attributes on VF vibration properties. For example, glottal halfwidth, vibration amplitudes, and time-constant phase differences likely relate to breathiness and brightness of the voice timbre. In particular, the level of additive noise increases with glottal area and flow, and relates to breathiness (Fraj et al., 2012; Schoentgen et al., 2015). Also, the spectral centroid and slope increase with the maximum declination rates of the area and the flow, and relate to brightness (Palaparthi et al., 2019; Schubert & Wolfe, 2006; Titze, 2006). However, it is currently unknown what impact the proposed perturbation types may have on acoustic features or auditory perception. To shed some light on the latter, an important future step may be to develop a graphical user interface that enables controlling the kinematic parameters while synthesizing the voice audio in real time. This may enable the auditory exploration of the multidimensional space of voice quality in dependence on kinematic VF parameters.

Second, the model may be used for the development of a more complete and technically sound terminology of

VF vibration patterns. For example, the shown perturbation patterns are often subsumed as "irregularity," which is a rather unspecific attribute. Instead, our model may support the development of a VF-kinematics-based framework for a more specific description of voice quality.

Third, the proposed synthesis model may be used for advanced testing of methods for analyzing LHSVs. For example, the measurement of perturbations from clinical videos lacks a controllable ground truth.

Finally, the effects of the kinematic parameters on the visual perception of the videos may be studied. This may promote the clinical acceptance of LHSVs, because clinicians may use the synthesizer to develop an intuition of how the proposed control parameters relate to the perceived vibration patterns.

Acknowledgments

Philipp Aichinger received a grant from the Austrian Science Fund (KLI722-B30). S. Pravin Kumar received a grant from SSN Trust (SSN/IFFP/January2019/1-12/08). Jan G. Švec received support from the Technology Agency of the Czech Republic (TH04010422 "VKG. 3.0") and a grant from Palacký University in Olomouc (IGA_PrF_2021_017). This work was supported by the Austrian Science Fund (FWF): KLI 722-B30, the Technology Agency of the Czech Republic Project No. TH04010422 "VKG 3.0," and the SSN Trust, SSN/IFFP/January2019/1-12/08. The stipend for doctoral studies of H.L. was provided by the Palacký University Internal Grant IGA_PrF_2021_017.

References

Aichinger, P., & Pernkopf, F. (2021). Synthesis and analysis-by-synthesis of modulated diplophonic glottal area waveforms. *IEEE/ACM Transactions on Audio Speech and Language Processing, 29,* 914–926. https://doi.org/10.1109/TASLP.2021.3053387

Aichinger, P., Roesner, I., Leonhard, M., Denk-Linnert, D., Bigenzahn, W., & Schneider-Stickler, B. (2016). A database of laryngeal high-speed videos with simultaneous high-quality audio recordings of pathological and non-pathological voices. *Proceedings of the International Conference on Language Resources and Evaluation (LREC), 10,* 767–770. https://doi.org/10.13140/RG.2.2.15467.34088

Berry, D., Montequin, D. W., Clark, M. J. O., & Titze, I. (2001). Characterization of the medial surface of the vocal folds. *Annals of Otology, Rhinology and Laryngology, 110*(5), 470–477. https://doi.org/10.1177/000348940111000514

Boessenecker, A., Berry, D., Lohscheller, J., Eysholdt, U., & Doellinger, M. (2007). Mucosal wave properties of a human vocal fold. *Acta Acustica united with Acustica, 93*(5), 815–823.

Bulusu, S., Kumar, S., Švec, J. G., & Aichinger, P. (2021). Fitting synthetic to clinical kymographic images for deriving kinematic vocal fold parameters: Application to left–right vibratory phase differences. *Biomedical Signal Processing and Control, 63,* 102253. https://doi.org/10.1016/j.bspc.2020.102253

Childers, D. G., Alsaka, Y. A., Hicks, D. M., & Moore, G. P. (1986). Vocal fold vibrations in dysphonia: Model vs. measurement. *Journal of Phonetics, 14*(3–4), 429–434. https://doi.org/10.1016/s0095-4470(19)30684-9

Cranen, B., & Schroeter, J. (1996). Physiologically motivated modelling of the voice source in articulatory analysis/synthesis. *Speech Communication, 19*(1), 1–19. https://doi.org/10.1016/0167-6393(96)00016-7

Döllinger, M., Berry, D., & Kniesburges, S. (2016). Dynamic vocal fold parameters with changing adduction in ex-vivo hemilarynx experiments. *The Journal of the Acoustical Society of America, 139*(5), 2372–2385. https://doi.org/10.1121/1.4947044

Fraj, S., Schoentgen, J., & Grenez, F. (2012). Development and perceptual assessment of a synthesizer of disordered voices. *The Journal of the Acoustical Society of America, 132*(4), 2603–2615. https://doi.org/10.1121/1.4751536

Herbst, C., Hampala, V., Garcia, M., Hofer, R., & Svec, J. G. (2017). Hemi-laryngeal setup for studying vocal fold vibration in three dimensions. *Journal of Visualized Experiments, 2017*(129), 1–9. https://doi.org/10.3791/55303

Jiang, J., Chang, C. I. B., Raviv, J. R., Gupta, S., Banzali, F. M., & Hanson, D. (2000). Quantitative study of mucosal wave via videokymography in canine larynges. *The Laryngoscope, 110*(9), 1567–1573. https://doi.org/10.1097/00005537-200009000-00032

Kumar, S., & Švec, J. G. (2019). Kinematic model for simulating mucosal wave phenomena on vocal folds. *Biomedical Signal Processing and Control, 49,* 328–337. https://doi.org/10.1016/j.bspc.2018.12.002

Lohscheller, J., & Eysholdt, U. (2008). Phonovibrogram visualization of entire vocal fold dynamics. *The Laryngoscope, 118*(4), 753–758. https://doi.org/10.1097/MLG.0b013e318161f9e1

Lohscheller, J., Eysholdt, U., Toy, H., & Dollinger, M. (2008). Phonovibrography: Mapping high-speed movies of vocal fold vibrations into 2-D diagrams for visualizing and analyzing the underlying laryngeal dynamics. *IEEE Transactions on Medical Imaging, 27*(3), 300–309. https://doi.org/10.1109/TMI.2007.903690

Lucero, J. C., Schoentgen, J., Haas, J., Luizard, P., & Pelorson, X. (2015). Self-entrainment of the right and left vocal fold oscillators. *The Journal of the Acoustical Society of America, 137*(4), 2036–2046. https://doi.org/10.1121/1.4916601

Mirza, M., & Osindero, S. (2014). Conditional generative adversarial nets. 1–7. http://arxiv.org/abs/1411.1784

Orlikoff, R., Golla, M. E., & Deliyski, D. (2012). Analysis of longitudinal phase differences in vocal-fold vibration using synchronous high-speed videoendoscopy and electroglottography. *Journal of Voice, 26*(6), 816.e13–816.e20.

Palaparthi, A., Smith, S., & Titze, I. R. (2019). Mapping thyroarytenoid and cricothyroid activations to postural and acoustic features in a fiber-gel model of the vocal folds. *Applied Sciences (Switzerland), 9*(21). https://doi.org/10.3390/app9214671

Samlan, R. A., & Story, B. H. (2017). Influence of left–right asymmetries on voice quality in simulated paramedian vocal fold paralysis. *Journal of Speech, Language, and Hearing Research, 60*(2), 306–321. https://doi.org/10.1044/2016_JSLHR-S-16-0076

Scherer, R. C., Shinwari, D., De Witt, K. J., Zhang, C., Kucinschi, B. R., & Afjeh, A. A. (2001). Intraglottal pressure profiles for a symmetric and oblique glottis with a divergence angle of 10 degrees. *The Journal of the Acoustical*

Society of America, 109(4), 1616–1630. https://doi.org/10.1121/1.1333420

Schoentgen, J., Fraj, S., & Lucero, J. C. (2015). Testing the reliability of grade, roughness and breathiness scores by means of synthetic speech stimuli. *Logopedics Phoniatrics Vocology, 40*(1), 5–13. https://doi.org/10.3109/14015439.2013.837502

Schubert, E., & Wolfe, J. (2006). Does Timbral Brightness Scale with frequency and spectral centroid? *Acta Acustica united with Acustica, 92*(5), 820–825.

Shiba, T. L., & Chhetri, D. K. (2016). Dynamics of phonatory posturing at phonation onset. *The Laryngoscope, 126*(8), 1837–1843. https://doi.org/10.1002/lary.25816

Šidlof, P., Švec, J. G., **Horáček, J., Veselý, J., Klepáček, I., & Havlík, R.** (2008). Geometry of human vocal folds and glottal channel for mathematical and biomechanical modeling of voice production. *Journal of Biomechanics, 41*(5), 985–995. https://doi.org/10.1016/j.jbiomech.2007.12.016

Story, B. H. (2013). Phrase-level speech simulation with an airway modulation model of speech production. *Computer Speech and Language, 27*(4), 989–1010. https://doi.org/10.1016/j.csl.2012.10.005

Titze, I. R. (1984). Parameterization of the glottal area, glottal flow, and vocal fold contact area. *The Journal of the Acoustical Society of America, 75*(2), 570–580. https://doi.org/10.1121/1.390530

Titze, I. R. (1988). The physics of small-amplitude oscillation of the vocal folds. *The Journal of the Acoustical Society of America, 83*(4), 1536–1552. http://scitation.aip.org/content/asa/journal/jasa/83/4/10.1121/1.395910

Titze, I. R. (2006). Theoretical analysis of maximum flow declination rate versus maximum area declination rate in phonation. *Journal of Speech, Language, and Hearing Research, 49*(2), 439–447. https://doi.org/10.1044/1092-4388(2006/034)

Titze, I. R., & Alipour, F. (2006). *The myoelastic aerodynamic theory of phonation*. National Center for Voice and Speech.

Vahabzadeh-Hagh, A. M., Zhang, Z., & Chhetri, D. K. (2017). Quantitative evaluation of the in vivo vocal fold medial surface shape. *Journal of Voice, 31*(4), 513.e15–513.e23. https://doi.org/10.1016/j.jvoice.2016.12.004

Verdonck-de Leeuw, I. M., Festen, J. M., & Mahieu, H. F. (2001). Deviant vocal fold vibration as observed during videokymography: The effect on voice quality. *Journal of Voice, 15*(3), 313–322. https://doi.org/10.1016/S0892-1997(01)00033-9

Wu, L., & Zhang, Z. (2016). A parametric vocal fold model based on magnetic resonance imaging. *The Journal of the Acoustical Society of America, 140*(2), EL159–EL165. https://doi.org/10.1121/1.4959599

Wu, L., & Zhang, Z. (2019). Voice production in a MRI-based subject-specific vocal fold model with parametrically controlled medial surface shape. *The Journal of the Acoustical Society of America, 146*(6), 4190–4198. https://doi.org/10.1121/1.5134784

Yokonishi, H., Imagawa, H., Sakakibara, K. I., Yamauchi, A., Nito, T., Yamasoba, T., & Tayama, N. (2016). Relationship of various open quotients with acoustic property, phonation types, fundamental frequency, and intensity. *Journal of Voice, 30*(2), 145–157. https://doi.org/10.1016/j.jvoice.2015.01.009

Department of Experimental Physics
Faculty of Science
Palacký University Olomouc, Czechia

Biomechanics and acoustics of voice production

Summary of Doctoral Thesis, May 2023

Author: Hugo Lehoux, MSc

Study program: Biophysics

Supervisor: Doc. RNDr. Jan G. Švec, Ph.D. et Ph.D.

Biomechanika a akustika tvorby hlasu / *Biomechanics and acoustics of voice production*

Uchazeč/*Applicant*: **Hugo Lehoux, M.Sc.**

Školitel/*Supervisor*: Doc. RNDr. Jan G. Švec, Ph.D. et Ph.D.

Univerzita Palackého, Přírodovědecká fakulta, Katedra biofyziky/*Palacký University, Faculty of Science, Department of Biophysics, Olomouc, Czech Republic*

Forma studia/*Form of study:* Prezenční/*Fulltime study*

Autoreferát byl rozeslán dne/ The thesis summary was distributed on: .....

Obhajoba se koná dne..... před komisí pro obhajoby disertačních prací v oboru Biofyzika v .....

The defence will occur on .....in front of the Committee for Defences of Ph.D. Theses of the Field of Biophysics in.....

S disertačni prací je možné se seznámit na studijním oddělení Přírodovědecké fakulty UP v Olomouci.

The Ph.D. thesis can be seen at the Study Department of the Faculty of Science at the Palacký University in Olomouc.

Table of contents

LIST OF PUBLICATIONS IN THIS THESIS

I. Lehoux, H., Hampala, V., & Švec, J. G. (2021). Subglottal pressure oscillations in anechoic and resonant conditions and their influence on excised larynx phonations. *Scientific Reports, 11*(1), 28. doi:10.1038/s41598-020-79265-3
Cited in: Švec et al. 2021; Lester-Smith et al. 2021; Herbst et al. 2023;

Authors' contribution: H.L.: formal analysis, investigation, data curation, writing—original draft, visualization. V.H.: methodology, subglottal tract model development, investigation. J.G.Š.: conceptualization, methodology, resources, writing—review and editing, supervision, funding acquisition.

Received **3rd place in the Student Scientific competition for the Dean's Award – 2020**) at Přírodovědecká fakulta UP, 17. listopadu 12, Olomouc, in the Physics section under doctoral division.

The work from this publication was presented by the author during the following events:

- Scientific Day "Acoustique de la voix et de la parole : modèles, mesures, maquettes [Acoustics of voice and speech: models, measurements, testbeds]" organized by the Speech Acoustics Group (GAP) of the French Acoustical Society (SFA) at the Institut de recherche et coordination acoustique/musique (IRCAM), Paris, France, 15/03/2023 - *Biomechanics and acoustics of vocal production: excised larynges, vocal registers and kinematic modelling* (Invited lecture, 30 minutes)
- Seminar at the Royal Institute of Technology (KTH), Stockholm, Sweden, 05/11/2021 - Ongoing and past activities at the Voice Research Lab, Palacký University, Olomouc, Czechia & Investigation of electroglottographic waveshapes using a kinematic model of the vocal fold vibrations (Lecture, 30 minutes)
- Czech Acoustical Society student seminar, 22/01/2021 (online) - Subglottal pressure oscillations in anechoic and resonant conditions and their influence on excised larynx phonations (Lecture, 15 minutes)
- e-Forum Acusticum, 7-11/12/2020 (online) - Development and Use of an Anechoic Subglottal Tract for Excised Larynx Experiments (poster presentation)
- 11th International workshop on Model and Analysis of Vocal Emissions for Biomedical Applications (MAVEBA), 17-19/12/2019 in Florence, Italy - Development and Use of an Anechoic Subglottal Tract for Excised Larynx Experiments. (Lecture, 15 minutes)

II. Lehoux, H., Herbst, C. T., Dobiáš, M., & Švec, J. G. (2023). Frequency jumps in excised larynges in anechoic conditions: A pilot study. *Journal of Sound and Vibration, 551,* 117607. doi: 10.1016/j.jsv.2023.117607
Authors' contribution: H.L.: Methodology, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization. CTH: Conceptualization, Investigation, Writing – review & editing. MD: Resources, Writing – review & editing. JGS: Conceptualization, Methodology, Investigation, Writing – review & editing, Supervision, Funding acquisition.

Received **3rd place in the Student Scientific competition for the Dean's Award – 2023**) at Přírodovědecká fakulta UP, 17. listopadu 12, Olomouc, in the Physics section under doctoral division.

The work from this publication was presented by the author during the following events:

- Scientific Day “Acoustique de la voix et de la parole : modèles, mesures, maquettes [Acoustics of voice and speech: models, measurements, testbeds]” organized by the Speech Acoustics Group (GAP) of the French Acoustical Society (SFA) at the Institut de recherche et coordination acoustique/musique (IRCAM), Paris, France, 15/03/2023 - *Biomechanics and acoustics of vocal production: excised larynges, vocal registers and kinematic modelling* (Invited lecture, 30 minutes)
- Czech Acoustical Society student seminar, Czech Technical University in Prague (ČVUT), Czechia, 19/01/2023 - Frequency jumps in excised larynges in anechoic conditions: A pilot study (Lecture, 15 minutes)

III. Lehoux, H., Popeil, L., & Švec, J. G. (2022). Laryngeal and Acoustic Analysis of Chest and Head Registers Extended Across a Three-Octave Range: A Case Study. *Journal of Voice (In press, early online)*. doi:10.1016/j.jvoice.2022.02.014
Cited in: Zita et al. 2022.

Authors’ contribution: H.L.: formal analysis, investigation, data curation, writing —original draft, visualization. L.P.: conceptualization, data acquisition, writing—review and editing. J.G.Š.: conceptualization, methodology, data acquisition, resources, writing—review and editing, supervision, funding acquisition.

Received **2nd place in the Student Scientific competition for the Dean's Award – 2021**) at Přírodovědecká fakulta UP, 17. listopadu 12, Olomouc, in the Physics section under doctoral division.

The work from this publication was presented by the author during the following events:

- Scientific Day “Acoustique de la voix et de la parole : modèles, mesures, maquettes [Acoustics of voice and speech: models, measurements, testbeds]” organized by the Speech Acoustics Group (GAP) of the French Acoustical Society (SFA) at the Institut de recherche et coordination acoustique/musique (IRCAM), Paris, France, 15/03/2023 - *Biomechanics and acoustics of vocal production: excised larynges, vocal registers and kinematic modelling* (Invited lecture, 30 minutes)
- 7th International Physiology and Acoustics of Singing Conference PAS7+, 05/05/2022 (online) - Laryngeal and acoustic register differences across the entire pitch range of a female professional singer (Lecture, 15 minutes)
- Czech Acoustical Society student seminar, 20/01/2022 (online) - Laryngeal and acoustic analysis of chest and head registers extended across a three-octave range: a case study (Lecture, 15 minutes)
- Seminar at the Royal Institute of Technology (KTH), Stockholm, Sweden, 05/11/2021 - Ongoing and past activities at the Voice Research Lab, Palacký University, Olomouc, Czechia & Investigation of electroglottographic waveshapes using a kinematic model of the vocal fold vibrations (Lecture, 30 minutes)
- CHOICE FOR VOICE: Crossing boundaries in voice, 3-5/09/2021 (online) - Singing Register Differences in Vocal Fold Oscillations Observed Across Three Octaves Through Laryngeal High-Speed Videoendoscopy (Poster presentation)
- Viennese Workshop on Voice Quality, 27/08/2021 (online) - Chest versus Head Register Differences in Vocal Fold Oscillations across the Pitch Range of Three Octaves (Lecture, 15 minutes)
- 12th International Conference on Voice Physiology and Biomechanics (ICVPB), 2-4/12/2020 (online) - Singing Register Differences in Vocal Fold Oscillations Observed Across Three Octaves Through Laryngeal High-Speed Videoendoscopy (Poster presentation)

IV. Aichinger, P., Kumar, S. P., Lehoux, H., & Švec, J. G. (2022). Simulated Laryngeal High-Speed Videos for the Study of Normal and Dysphonic Vocal Fold Vibration. *Journal of Speech Language and Hearing Research, 65*(7), 2431-2445. doi:10.1044/2022_jslhr-21-00673

Authors' contribution: P.A.: conceptualization, software development, data acquisition, formal analysis, investigation, writing —original draft, visualization, funding acquisition. S.P.K.: software development, writing—review and editing. **H.L. software development, writing—review and editing**. J.G.Š.: conceptualization, writing—review and editing, supervision, funding acquisition.

ABSTRACT

Voice is a complex phenomenon, originating from the airflow modulated by the vibrations of the vocal folds, located in the larynx. The vibratory pattern of the vocal folds is influenced by the action of the intrinsic and extrinsic laryngeal muscles, which impacts the sound quality of the voice. The vocal folds also vibrate in a system comprising supraglottal and subglottal cavities presenting acoustic resonances, which can interact with the vocal fold vibrations. To better understand the mechanisms of voice production, the investigation of the vocal fold vibratory properties, and how they interact with the supraglottal and subglottal resonances, is therefore essential. This dissertation carries out this task by focusing on specific phenomena that are not sufficiently explained, such as the influence of the subglottal resonances on subglottal pressures and on the vocal fold vibrations, the occurrences of frequency jumps, and the notion of voice registers. This dissertation provides novel results, which help to improve our understanding of those phenomena, and can be divided into four main parts, related to four original articles.

In the first part, the development of an anechoic subglottal tract, to allow the observation of the vocal fold vibrations in subglottally anechoic conditions, is described. The acoustic response of the anechoic subglottal tract is first measured and compared with the acoustic response of a resonant subglottal tract, which reveals that the anechoic subglottal tract successfully suppresses its acoustic resonances. It is then used during excised deer larynx experiments without a supraglottal tract, where the subglottal pressure and the frequency of the vocal fold oscillation are measured, and compared with resonant conditions. In anechoic conditions, the subglottal pressure waveform resembles the theoretical source flow waveform, exhibiting a constant value during the closed phase, whereas it contains strong fluctuations during the closed phase in resonant conditions. Additionally, in resonant conditions the oscillation frequency of the vocal folds shows consistently lower values than in anechoic conditions. This indicates that the presence of subglottal resonances alters not only the subglottal pressures but also the vocal fold vibrations.

In the second part, the developed anechoic subglottal tract is further used to investigate frequency jumps in both anechoic and subglottally resonant conditions, using human excised larynges without a supraglottal tract. In both these conditions, larynges exhibit consistent frequency jumps, which suggests that supraglottal or subglottal resonances are not needed for the jumps to occur. This indicates that such jumps are primarily caused by inherent nonlinear-dynamic properties of the larynges. Nevertheless, some differences are observed in the jumps between anechoic and resonant conditions, indicating that the subglottal resonances also have an influence, although secondary.

In the third part, chest and head registers are investigated in vivo in a professional female singer using laryngeal high-speed videoendoscopy, electroglottography, and an acoustic analysis of a microphone signal of the radiated sound. Chest and head registers are commonly assumed to be limited to low frequencies and high frequencies, respectively, and are mostly studied within a limited frequency range where the production of both registers is comfortable. Here, the singer produced both chest and head registers across a wide frequency range. The aim of this study is to find relevant parameters of vocal fold vibrations that allow to discriminate between the two registers, regardless of frequency. The microphone sound is used in blind listening tests to determine whether the intended registers can be audibly distinguished and reveals that the register changes are not always perceivable by listeners. Kinematically, the singer consistently differed the chest register from the head register by means of a more rapid glottal closure. This was usually, but not necessarily, accompanied by a greater membranous adduction of the vocal folds.

Finally, a kinematic model of the vocal fold vibrations for generating synthetic kymograms resembling those observed in vivo is extended to three dimensions, to allow the generation of synthetic high-speed videos of the entire glottis. This aims at helping to improve our interpretation of laryngeal high-speed videoendoscopy, especially in a clinical context where this technique is not yet frequently used. Synthetic high-speed videos corresponding to different types of phonation are generated. Specifically, the model is capable of generating videos exhibiting both regular vibrations, which are typical for

healthy voices, and irregular vibrations, which are usually associated with voice pathologies. The extended model can be used to study the influence of the geometric and kinematic VF parameter variation on the appearance of the vibratory patterns, as observed through high-speed videos, kymograms, and phonovibrograms.

ABSTRAKT

Hlas je složitý jev, který vzniká prouděním vzduchu modulovaným vibracemi hlasivek umístěných v hrtanu. Vibrace hlasivek jsou ovlivňovány činností vnitřních a vnějších hrtanových svalů, což má vliv na kvalitu hlasu. Navíc, hlasivky kmitají v traktu zahrnujícím supraglotické a subglotické dutiny, které vykazují akustické rezonance, jež mohou interagovat s vibracemi hlasivek. Pro lepší pochopení mechanismů tvorby hlasu je proto nezbytné zkoumat vibrační vlastnosti hlasivek a jejich interakci se supraglotickými a subglotickými rezonancemi. Tato disertační práce plní tento úkol tím, že se zaměřuje na specifické jevy, které nejsou dostatečně vysvětleny, jako je vliv subglotických rezonancí na kmity hlasivek, na přeskoky frekvence hlasu a na hlasové rejstříky. Tato disertační práce přináší nové výsledky, které přispívají k lepšímu pochopení těchto jevů, a lze ji rozdělit do čtyř hlavních částí, které souvisejí se čtyřmi původními články.

V první části je popsán vývoj bezdozvukového subglotického traktu, který odstraňuje akustické rezonance a umožňuje pozorování vibrací hlasivek v bezdozvukových podmínkách. Akustická odezva bezdozvukového subglotického traktu je nejprve změřena a porovnána s akustickou odezvou rezonančního subglotického traktu. Výsledky ukazují, že vyvinutý trakt úspěšně potlačuje své akustické rezonance. Poté je tento trakt použit při experimentech s preparáty hrtanu, kde jsou subglotický tlak a frekvence kmitání hlasivek porovnávány mezi bezdozvukovými a rezonančními podmínkami. V bezdozvukových podmínkách se průběh subglotického tlaku podobá teoretickému průběhu zdroje hlasu a vykazuje konstantní hodnotu během fáze uzavření glotis, zatímco v rezonančních podmínkách obsahuje silné fluktuace během fáze uzavření glotis. V rezonančních podmínkách navíc frekvence kmitání hlasivek vykazuje konzistentně nižší hodnoty než v bezdozvukových podmínkách, což ukazuje, že přítomnost subglotických rezonancí mění nejen subglotický tlak, ale i kmitání hlasivek.

Ve druhé části je vyvinutý bezdozvukový subglotický trakt použit ke zkoumání přeskoků frekvence hlasu jak v bezdozvukových, tak v subgloticky rezonančních podmínkách, a to za použití lidských preparátů hrtanu, bez připojeného supraglotického traktu. V obou podmínkách vykazují hrtany konzistentně přeskoky frekvence kmitání hlasivek, což dokazuje, že supraglotické nebo subglotické rezonance nejsou pro vznik přeskoků nutné. To naznačuje, že tyto přeskoky frekvenci jsou primárně způsobeny inherentními nelineárně-dynamickými vlastnostmi hrtanu. Nicméně byly pozorovány i drobné rozdíly v přeskocích frekvencí mezi bezdozvukovými a rezonančními podmínkami, což naznačuje, že subglotické rezonance zde mají také vliv, i když druhotný.

Ve třetí části je studován hrudní a hlavový rejstřík u profesionální zpěvačky pomocí vysokorychlostní videoendoskopie hrtanu, elektroglotografie a akustické analýzy signálu z mikrofonu registrujícího vyzařovaný zvuk. Běžně se předpokládá, že hrudní a hlavový rejstřík je omezen na nízké, respektive vysoké frekvence, a tyto rejstříky se studují v omezeném frekvenčním rozsahu, kde lze pohodlně produkovat oba rejstříky. V našem případě zpěvačka produkovala jak hrudní, tak hlavový rejstřík v širokém frekvenčním rozsahu. Cílem této studie bylo nalézt relevantní parametry kmitání hlasivek, které umožňují rozlišit oba rejstříky bez ohledu na frekvenci. Zvuk z mikrofonu byl použit při zaslepených poslechových testech, pro zjištění, zda lze oba rejstříky poslechově rozlišit. Testy odhalily, že změny rejstříků nejsou posluchači vždy dobře rozpoznatelné. Kinematicky se hrudní rejstřík odlišoval od hlavového rychlejším uzavíráním glotis. To bylo obvykle, ale ne nutně, doprovázeno výraznější addukcí blanité části glottis.

V poslední části je vylepšen kinematický dvojrozměrný model kmitání hlasivek pro generování syntetických kymogramů, které se podobají kymogramům pozorovaným in vivo. Model je rozšířen do tří rozměrů, aby bylo možné generovat syntetické vysokorychlostní videozáznamy celé glottis. Cílem je pomoci zlepšit interpretaci záznamů z vysokorychlostní videoendoskopie hrtanu, zejména v klinickém kontextu, kde je tato technika využívána pouze zřídka. Pomocí modelu jsou generována syntetická vysokorychlostní videa odpovídající různým typům fonace. Je ukázáno, že model je schopen generovat videa vykazující jak pravidelné kmity, které jsou typické pro zdravé hlasy, tak nepravidelné kmity, které jsou obvykle spojeny s patologiemi hlasu. Rozšířený model lze použít ke studiu vlivu geometrických a kinematických změn

parametrů VF na průběh kmitů, tak jak jsou pozorovány prostřednictvím vysokorychlostních videí, kymogramů a fonovibrogramů.

1. INTRODUCTION

1.1. Fundamentals/basics of voice production

Voice production is a complex phenomenon which normally involves many organs in several parts of the body. Breathing muscles (intercostal muscles, abdominal muscles, diaphragm, etc.) create an overpressure in the lungs below the vocal folds (VFs – see Figure 1), which generates the power needed for voice production. This power is transmitted to the VFs in the larynx and causes them to oscillate, which is a process called phonation. The sound produced by the VF vibrations propagates below (in the trachea) and above (in the pharynx, oral cavity, and nasal cavity) the VFs, and is radiated through the lips and sometimes through the nostrils, when the velum opens (see Figure 1). The oral and nasal cavities act as filters on the sound generated by the VF vibrations, which modifies its timbre in order to create different vowels, for example. Speech commonly uses sounds which are either voiced (when the VFs vibrate) or unvoiced (where the VFs do not vibrate), however this thesis only focuses on the vibration of the VFs, and therefore will not discuss unvoiced sounds.

The space between the VFs is referred to as the glottis (see Figure 2), and therefore the cavities below and above the glottis are referred to as the subglottal (below the glottis) tract and the supraglottal (above the glottis) tract (also sometimes called the vocal tract). Figure 1 shows a schematic diagram of various organs used in voice production.

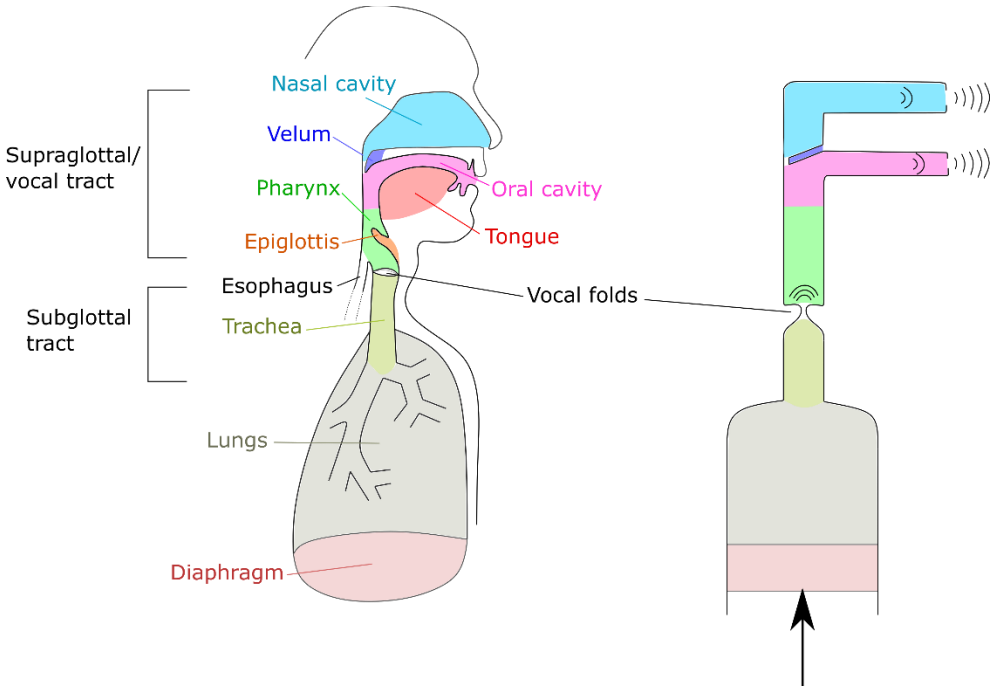


Figure 1. Basic anatomy of the phonatory system. Modified from Flanagan (1972).

1.1.1. Anatomy of the vocal apparatus

1.1.1.1. Larynx

The VFs are attached anteriorly to the thyroid cartilage, and posteriorly to the arytenoids (see Figure 3a). The arytenoids are attached to the superior and posterior ridge of the cricoid cartilage (see Figure 3a). The thyroid cartilage is also attached to the cricoid cartilage through the cricothyroid articular joint (see Figure 3b), which allows the thyroid cartilage to rotate anteriorly and elongate the VFs through the action of the cricothyroid (CT) muscle (see Figure 3). The cricothyroid joint is not strictly a rotational axis, as there can also be some small translational (anterior-posterior) component to the

movement, however this translational component has been shown to have a smaller influence than the rotational component (Vilkman, 1987), therefore we do not consider the translational component in this thesis.

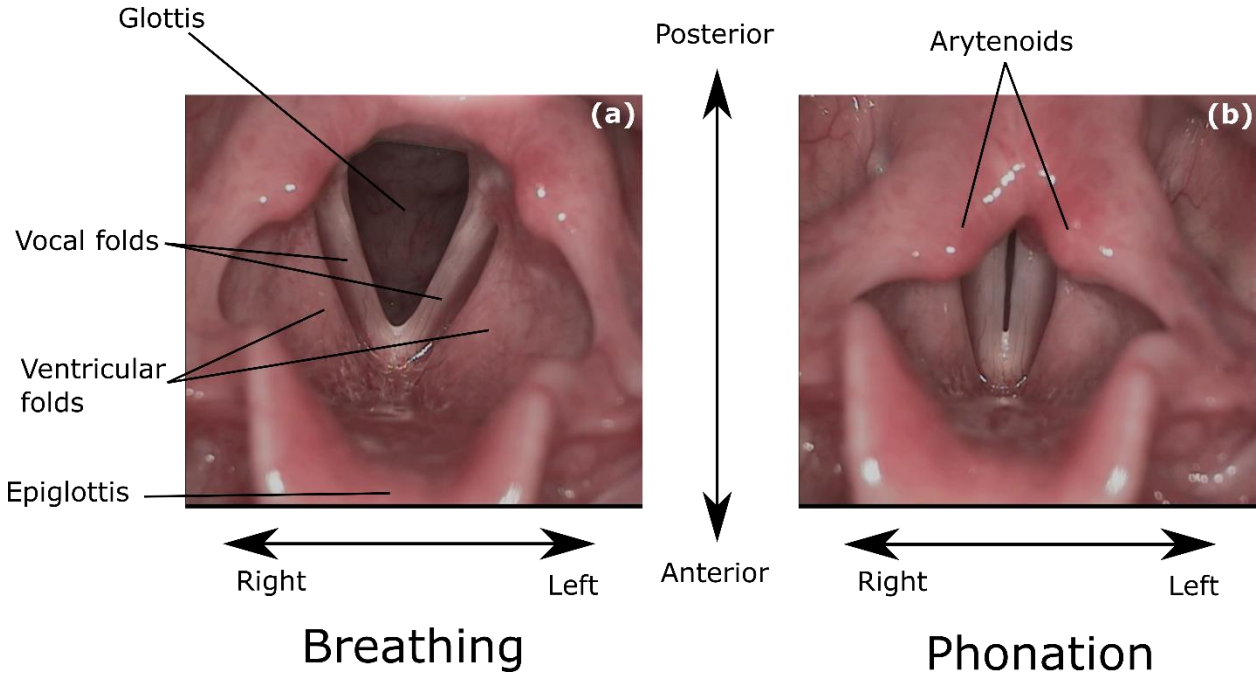


Figure 2. Laryngoscopic images showing the anatomy of the larynx. (a) During breathing; (b) During phonation. Images taken from laryngeal videos, courtesy of MUDr. Jitka Vydrová, Voice Centre Prague, Medical Healthcom Ltd, Czechia.

The intrinsic laryngeal muscles described in Figure 4 are used to move and rotate the arytenoids with respect to the cricoid cartilage. The action of those muscles either leads to the abduction (parting away from the midline) or the adduction (approximation towards the midline) of the VFs. The abductor muscle is the posterior cricoarytenoid (PCA) muscle (Figure 4b), and the adductor muscles are the lateral cricoarytenoid (LCA) muscle (Figure 4c) and the interarytenoid (IA) muscle (Figure 4d) (Van den Berg et al., 1960; Titze, 2000a).

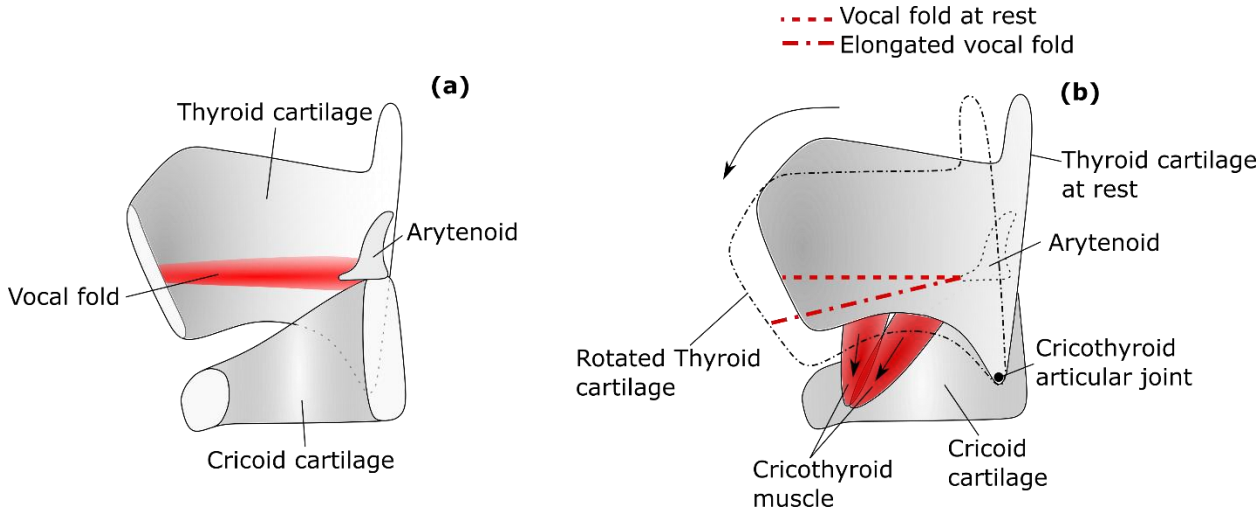


Figure 3. (a) Description of the laryngeal cartilages and the vocal fold attachment; (b) Description of the thyroid cartilage rotation and the vocal fold elongation, as a result of the CT muscle action. Modified from Titze et al. (1988), Fig. 1.

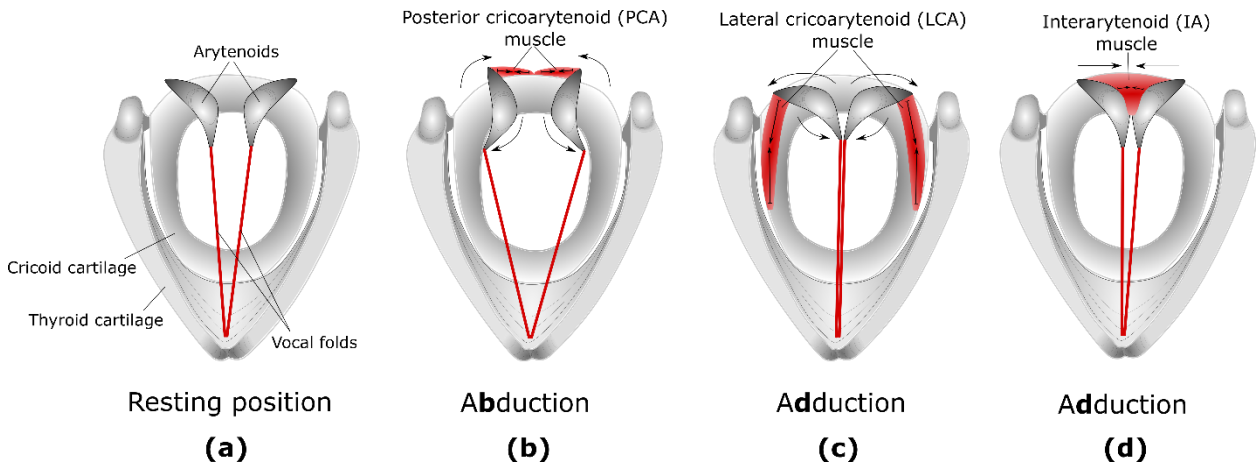


Figure 4. Description of the intrinsic laryngeal muscles used for abduction and adduction. Modified from Zemlin (1981).

VFs are complex tissues composed of five different layers (Hirano, 1974, 1975; Titze, 2000a): the epithelium, the superficial, intermediate, and deep layer of the lamina propria, and the thyroarytenoid (TA) muscle layer (see Figure 5). The layers have different compositions and therefore different biomechanical properties. They are sometimes grouped together into two main parts called the "body" (comprising the muscle layer and the deep layer of the lamina propria) and the "cover" (comprising the superficial and intermediate layers of the lamina propria, and the epithelium) (Hirano, 1974, 1975; Titze et al., 1988; Titze, 2000a).

Because the VFs are mainly composed of the TA muscle, their shape and biomechanical properties vary depending on the activity of this muscle. The contraction of the TA muscle will tend to actively shorten, tense, and bulge the VFs (see Figure 6) (Hirano, 1974, 1975; Wu et al., 2019). Alternatively, the action of the CT and LCA muscles will also passively tense the vocal fold tissues from their elongation.

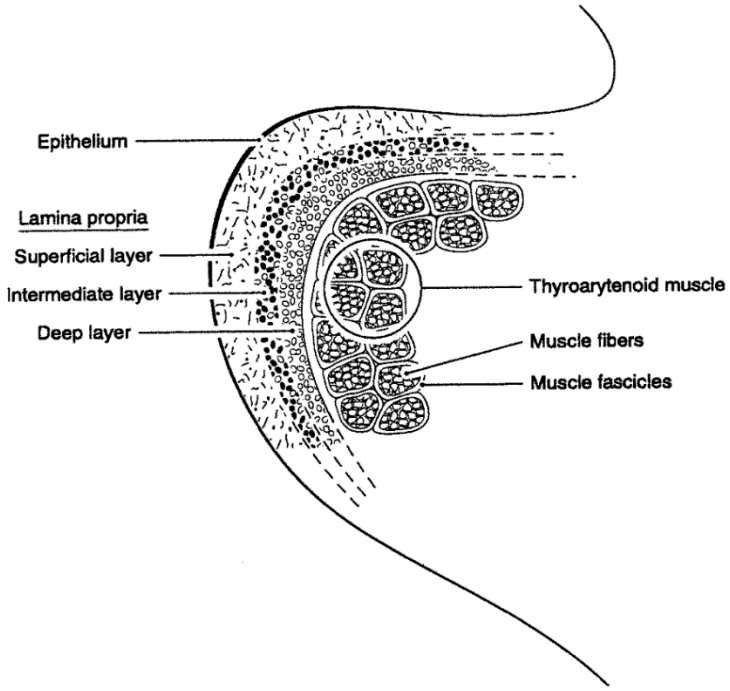


Figure 5. Coronal view of the VF and its different layers. Taken from Titze (2000a).

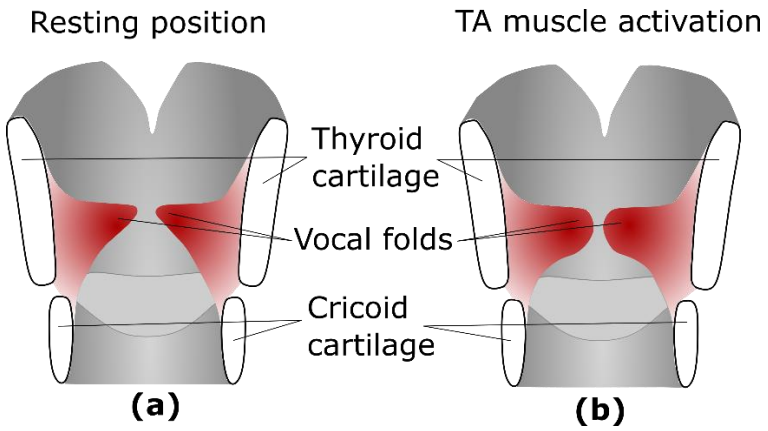


Figure 6. Effect of the TA muscle activation. (a) Resting position, the VFs are thin; (b) when the TA muscle activates, the VFs are bulged and thicker. Based on histological data from Hirano (1975).

1.1.1.2. Supraglottal cavities

The supraglottal tract (also called vocal tract) has an average effective length of about 17 cm (Baer et al., 1991; Story et al., 1996; Fitch et al., 1999; Vorperian et al., 2009), but its effective length can be increased or decreased by moving the larynx up (for example when swallowing) and down (for example when yawning). The supraglottal tract can be assimilated to a curved tube, through which the sound generated by the VF vibrations propagates and reflects at the mouth opening (as indicated on Figure 1). The sound can also propagate through the nasal cavity and reflect at the nostrils, but often the entrance to the nasal cavity is closed by the velum and the nasal cavity can then be disregarded. Constrictions in the supraglottal tract, caused by e.g. the tongue placement, can also cause the sound to reflect. Therefore, the acoustic transfer function of the supraglottal tract exhibits local maxima, which are associated with acoustic resonances. The frequency of those resonances depend on the geometry of the supraglottal tract, particularly on its effective length and the position of its constrictions. Figure 7 depicts the supraglottal tract geometry and the corresponding acoustic transfer function, for two different English vowels.

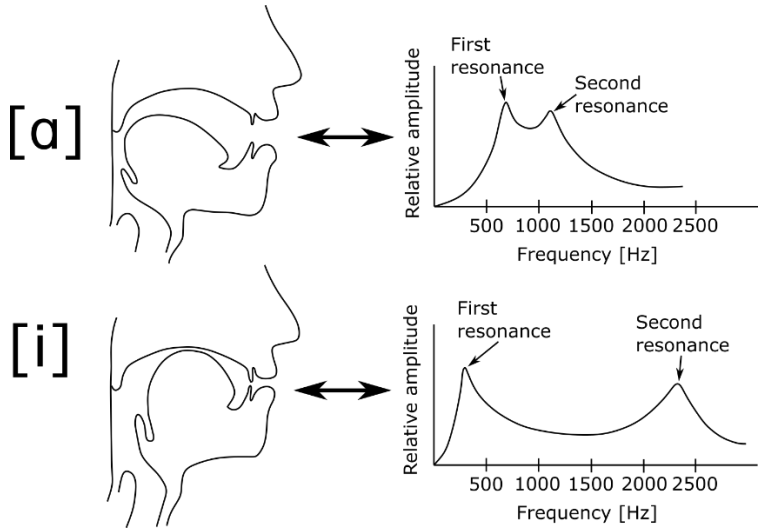


Figure 7. Description of the changes in the acoustic properties of the supraglottal tract caused by changes in its geometry. The two examples show the supraglottal tract geometry and the corresponding transfer function for the [a] and the [i] vowels. Modified from Titze (2000c).

1.1.1.3. Subglottal cavities

The subglottal tract consists of the cavity below the glottis, down to the point where the trachea splits into the bronchi. The length of the trachea, as measured from the lower border of the cricoid cartilage down to the inferior border of the carina (ridge of cartilage at the base of the trachea, separating the openings of the left and right main bronchi – see Figure 8) is, on average, about 12 cm (Grillo et al., 1964). This length can increase by up to 2 cm with deep inspiration (Holbert et al., 1995).

The trachea is rather straight and presents acoustic resonances f_{Rn} which approximately correspond to those of a straight tube open at one end and closed at the other end. Frequencies of such acoustic resonances can be expressed by the equation $f_{Rn} \approx (2n - 1)f_{R1}$, where n is an odd integer. Cranen et al. (1987) measured average values of f_{R1} = 510 Hz and f_{R2} = 1355 Hz, while other studies found comparable values (Fant et al., 1972; Lulich et al., 2012; Sundberg et al., 2013). The soft walls of the trachea create a large acoustic damping due to viscous losses, which is reflected in the high bandwidth (>100 Hz, Q factor > 0.2 for the first resonance) of the subglottal resonances, as measured by e.g., Cranen et al. (1987) and Fant et al. (1972).

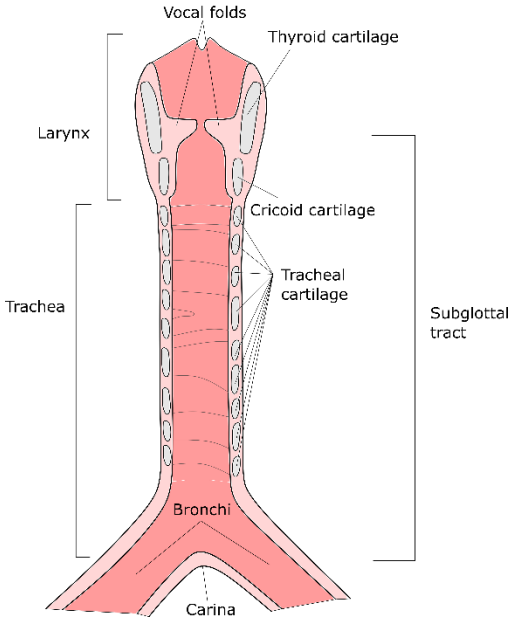


Figure 8. Simplified diagram of the larynx and the subglottal tract.

1.2. Voice instrumentation and parameterization

To better understand the mechanisms of voice production, researchers have designed methods to directly or indirectly visualize and capture the voice, particularly the VF vibrations. This section only describes the techniques, which were used in the papers published as parts of this thesis; other techniques exist but their description is beyond the scope of this thesis.

1.2.1. Laryngoscopic imaging and glottal parameters

1.2.1.1. Standard videolaryngoscopy

Laryngoscopy originates from the need for visually observing the VF vibrations in the larynx. The first instances of laryngoscopy were performed, as early as the middle of the 19th century, by placing a tilted mirror at the back of the mouth, and pointing a light source towards the mirror, in order to have a good visibility of the laryngeal tissues (Moore, 1991; Alberti, 1996; Baken et al., 2000). This simple idea created the basis of modern laryngoscopy, which commonly uses a tube called an endoscope. The endoscope can be either flexible (inserted through the nose, see Figure 9a) or rigid

(inserted through the mouth, see Figure 9b), and a camera is commonly connected to it, allowing to record and store the images for later playback and analysis.

Standard videolaryngoscopy can give qualitative insights on the laryngeal configurations and their variations in time, such as the adduction of the VF and the ventricular folds, or the position of the epiglottis and other surrounding tissues. It can also be used in a clinical context to observe the general state of the VFs and other surrounding tissues, and to discover pathologies such as polyps or other lesions (Hirano, 1981). However, the oscillations of the VFs are too fast to be observed and captured using standard videolaryngoscopy.

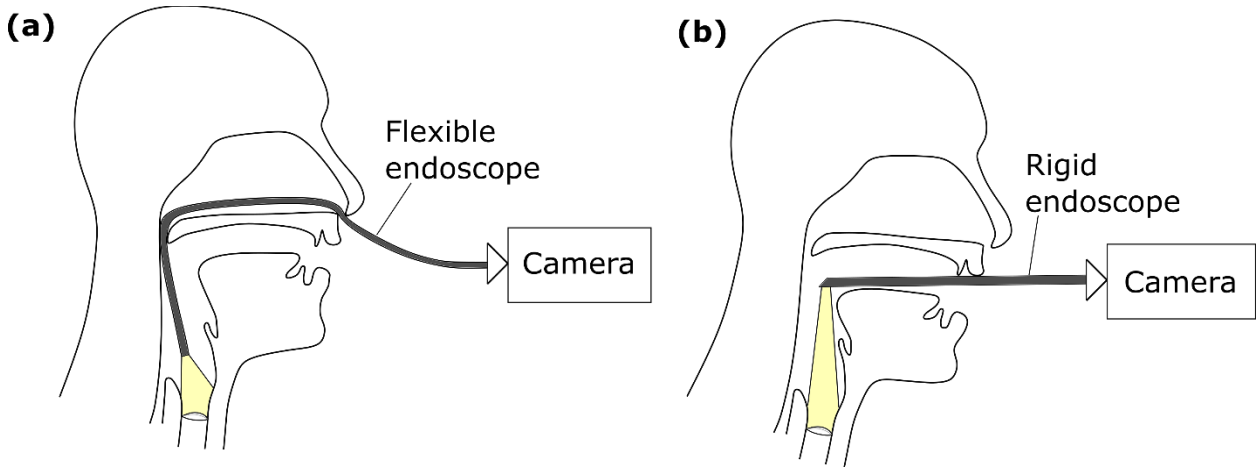


Figure 9. The two main types of endoscope used for laryngoscopic imaging. (a) flexible endoscope through the nasal cavity. (b) Rigid endoscope through the oral cavity.

1.2.1.2. Strobolaryngoscopy

In order to observe the oscillations of the VFs, a technique called strobolaryngoscopy was developed. This technique uses a standard camera with a stroboscopic light source, emitting a series of very short light flashes. The frequency of oscillations of the VFs is measured in real time from the electroglottographic (to be described in section 1.2.2) or the microphone signal, and the frequency of the strobe light is set to be slightly (commonly 1 to 2 Hz) lower than the frequency of oscillations., creating the illusion of a slow oscillation (Kallen, 1932; Moore, 1991; Verikas et al., 2009; Mehta et al., 2010). Strobolaryngoscopy is the most commonly used technique in clinical practice due to its direct, real-time interpretability, and it can be used to estimate parameters related to the VF vibrations (Casiano et al., 1992; Mehta et al., 2010). However, because of the nature of strobolaryngoscopy and its dependence on regular VF vibration with a steady oscillation frequency, one of its major limitations is the impossibility to properly detect irregular phonations.

1.2.1.3. Laryngeal high-speed videoendoscopy

To overcome the limitations of strobolaryngoscopy, researchers can use a high-speed video camera instead of a standard video camera. A high-speed camera can capture images at a much higher rate than the standard (25-50 frames per second – FPS) rate. Technological advances allow cameras to record several tens of thousands of FPS (see, e.g., Herbst, Lohscheller, et al. (2014)). Deliyski, Powell, et al. (2015) found that a frame rate of 8,000 FPS (with a minimum requirement of 4,000 FPS) is capable to observe the VF vibrations with sufficient detail. Laryngeal high-speed videoendoscopy (laryngeal HSV – see Deliyski, Hillman, et al. (2015) for a rationale behind the terminology) allows to observe phenomena which cannot be observed through strobolaryngoscopy, such as irregular vibrations, sudden register changes, or phonation onsets and offsets (Farnsworth, 1940; Eysholdt et al., 1996; Hertegård, 2005). Despite its superior potential for interpretation of the VF vibrations, laryngeal HSV is rarely used in clinical practice, because it does not allow direct, real-time visualization. Since the motion is registered at a very high frame rate, visualization of the frames is very time-consuming and is not suitable in a fast-pace clinical context with many patients.

1.2.1.4. Glottis segmentation and phonovibrography (PVG)

To facilitate the visualization of the VF vibrations captured in laryngeal HSV, image processing techniques can be used to segment the glottis, which means that the glottal area and edges are detected for each frame (Kist et al., 2021). Glottis segmentation allows to extract the time-varying glottal area in pixels (usually referred to as glottal area waveform – GAW) as well as the glottal midline (Kist et al., 2020) and the distance of the left and right VF edge from the glottal midline.

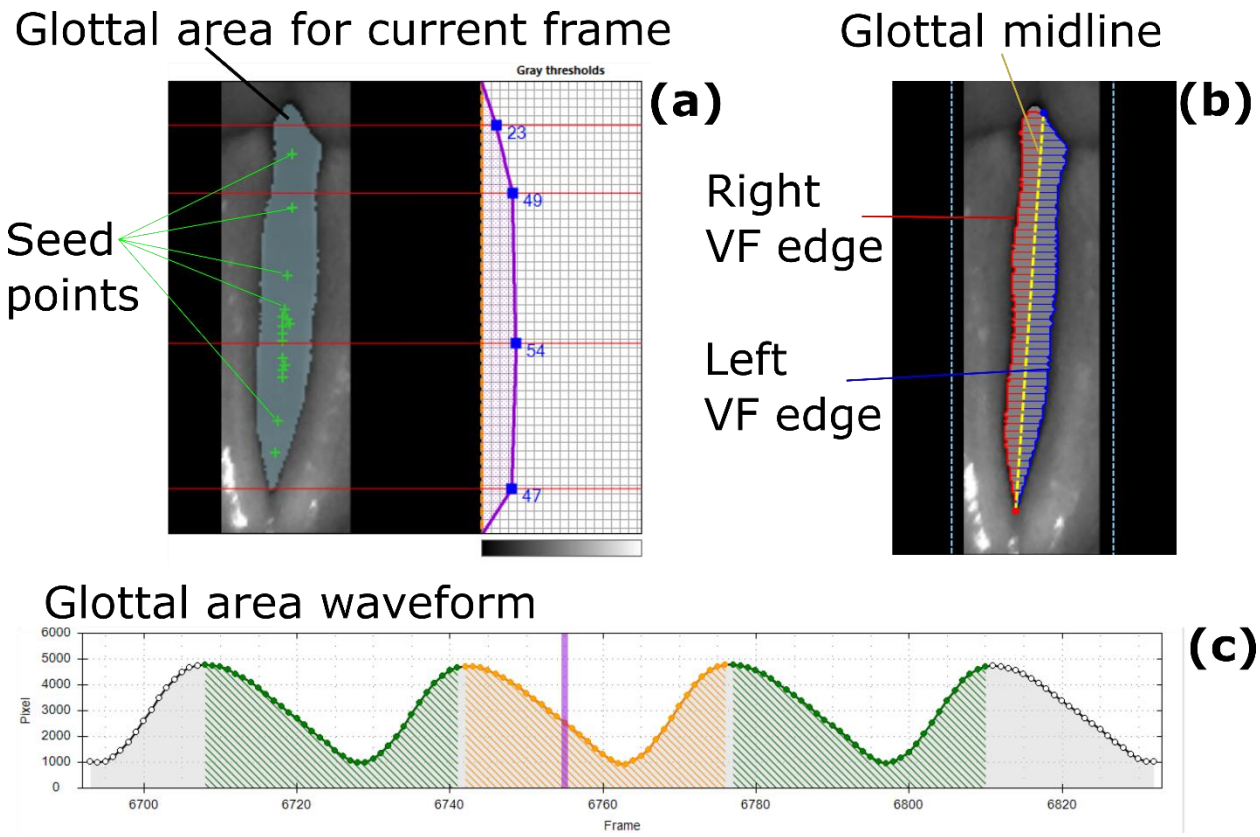


Figure 10. Description of the glottis segmentation from laryngeal HSV data, using Glottis Analysis Tools (Kist et al., 2021). (a) Seed points (green crosses) are placed inside the visible glottis, and the threshold of pixel intensity is adjusted at four different positions along glottal length, so that the detected glottal area (blueish-gray area) matches the visible glottis; (b) Once the segmentation parameters are adjusted to give a satisfactory glottis detection, the glottal midline (yellow dashed line) is automatically detected for each oscillatory cycle, and the distance of the VF edges from this line is coded for each frame; (c) The size of the glottal area (in pixels) is saved for each frame, and displayed as the glottal area waveform (GAW).

Another technique for more practical visualization and quantification of laryngeal HSV is phonovibrography (PVG – Figure 11). This technique consists in detecting the left and right glottal edges and saving their respective distances from the glottal midline (Figure 11a). Then, for each frame, the edge distances along the glottal midline are color-coded (Figure 11b) and concatenated to form the PVG image (Figure 11c). This technique has the benefit of allowing to quickly visualize the motion of the entire glottis, especially regarding opening and closing events (Lohscheller et al., 2008; Döllinger et al., 2009). Other visualization techniques exist, such as glottovibrography (Karakozoglou et al., 2012), but they were not used in the work presented in this thesis, and therefore their description is beyond the scope of this thesis.

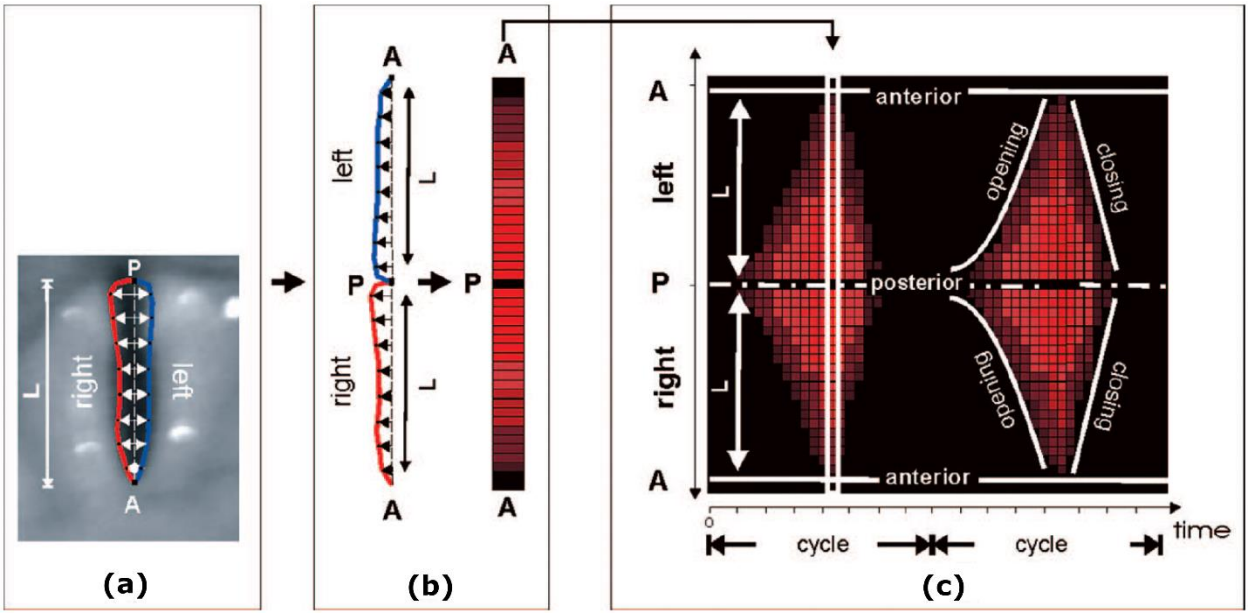


Figure 11. Description of the PVG generation. (a): the glottis is first segmented to detect the glottal edges, for the left and right VFs; (b) the edge contour for the left VF is flipped over the right contour, and the edge distances from the glottal midline are color-coded (here the red color indicates a larger distance, and the black color indicates a smaller distance); (c) the color-coded edge distances are then concatenated for every frame of the HSV. From Lohscheller et al. (2008).

1.2.1.5. Kymography

Kymography is a visualization method used to facilitate the interpretation of HSV images. Digital kymograms (DKG) are images obtained by considering a straight, steady scanline, and concatenating the line images corresponding to this scanline for each HSV frame. An example is shown on Figure 12, where the HSV frame is depicted on the left, showing the straight scanline that was used to extract the line images for every frame, concatenated to form the DKG image, shown on the right.

A technique called videokymography (VKG), developed by Švec et al. (1996), allows the visualization of kymograms in real time. This is done by a modified video camera scanning one horizontal line at a high rate (7,812.5 and 7,200 lines per second for the first and second generation of VKG cameras, respectively – see Qiu et al. (2007) for details about the second generation camera) and displaying the resulting kymographic images at 25 or 30 FPS.

Kymography and VKG have been extensively used to investigate normal and pathological vibrations (Švec et al., 2002; Švec et al., 2007, 2009; Phadke et al., 2017). Notably, VKG has been shown to be useful in assessing the quality of the VF mucosa, through evaluation of the sharpness of the lateral peaks (Jiang et al., 2000; Yamauchi et al., 2015; Kumar et al., 2020). The kymogram image can also be segmented to extract what the author refers to as the kymographic contour, using e.g., VKGanalyzer (Novozámský et al., 2015; Zita et al., 2022).

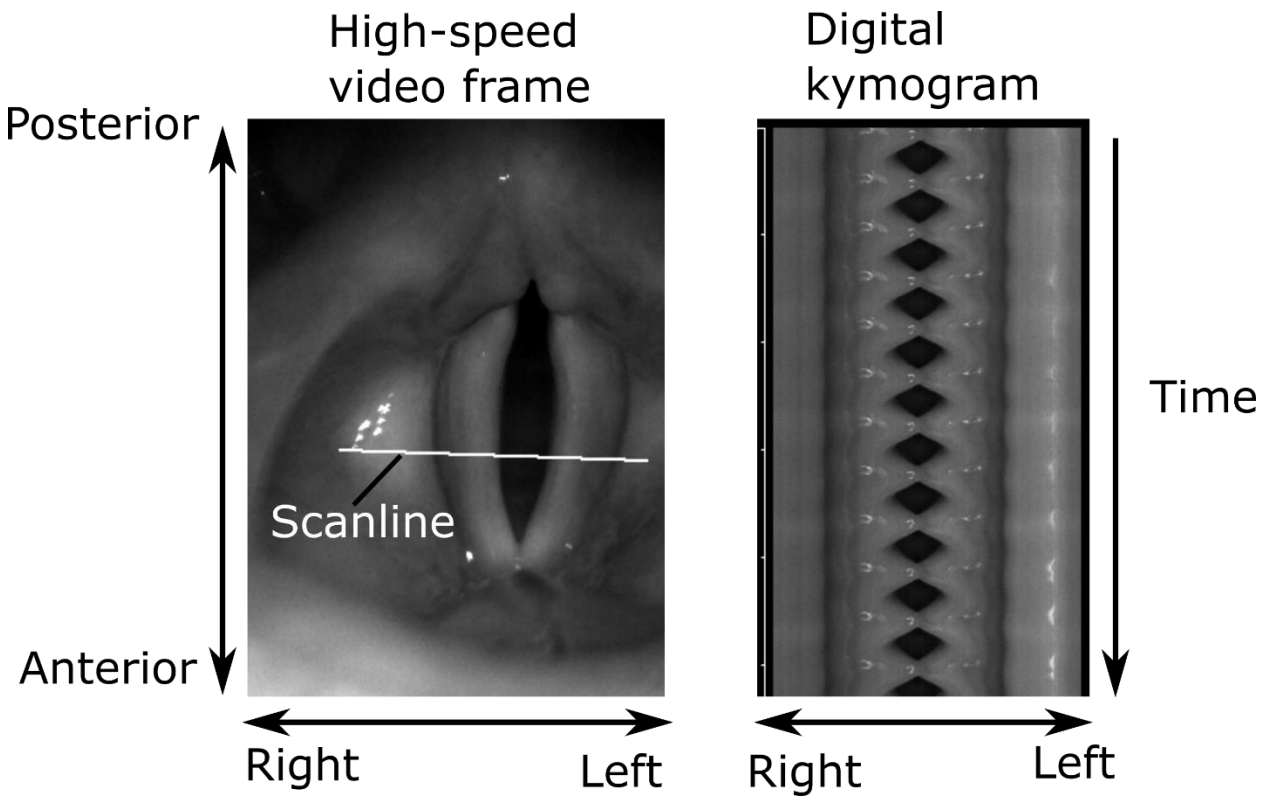


Figure 12. Example of a digital kymogram. The left panel shows a high-speed video frame with the manually drawn scanline. The right panel shows the kymogram image, which concatenates the pixels of the scanline vertically for each subsequent frame.

1.2.1.6. Analysis of the VF vibrations from laryngoscopic images

Segmentation of laryngeal images and kymograms can be used to obtain glottal contours and calculate parameters related to the VF vibrations, which provide a more objective analysis and offer the possibility of comparison and replication. Such parameters are commonly related to the durations of different phases during an oscillation cycle, and their relations to each other. Figure 13 gives a representation of those phases: the oscillation cycle is divided into the closed phase (when the glottis is closed) and the open phase (when the glottis is open), while the open phase is further subdivided into the opening and the closing phases. Here, we define their durations as t_{closed} (closed phase duration), t_{open} (open phase duration), $t_{closing}$ (closing phase duration), $t_{opening}$ (opening phase duration), and t_{cycle} (cycle duration). Their values can be measured from the GAW, the PVG (see section 1.2.1.4), or the kymographic contour (see section 1.2.1.5), and several metrics can be derived (Hirano, 1981). For brevity, only the parameters investigated in paper **III** of this thesis are described here (see e.g. Schlegel et al. (2019) for a more comprehensive list):

- The closed quotient (CQ) is the ratio of the closed phase duration and the cycle duration:

$$CQ = \frac{t_{closed}}{t_{cycle}}. \qquad \text{Eq. (1)}$$

This quotient is sometimes replaced by the open quotient (OQ), which is the ratio of the open phase duration and the cycle duration. The CQ and OQ are substitutable measures, as one can be calculated from the other: OQ = 1 - CQ.

- The speed quotient (SQ) is the ratio of the opening phase and the closing phase durations:

$$SQ = \frac{t_{opening}}{t_{closing}}. \qquad \text{Eq. (2)}$$

An alternative expression of the SQ is the speed index (SI), which is the difference between the opening and closing phase durations, divided by the sum of the opening and closing phase durations:

$$SI = \frac{t_{opening} - t_{closing}}{t_{opening} + t_{closing}}, \quad \text{Eq. (3)}$$

- The closing quotient (CgQ) and opening quotient (OgQ) are the ratios of the closing phase duration to the cycle duration, and of the opening phase duration to the cycle duration, respectively:

$$CgQ = \frac{t_{closing}}{t_{cycle}}, \quad \text{Eq. (4)}$$

$$OgQ = \frac{t_{opening}}{t_{cycle}}, \quad \text{Eq. (5)}$$

- The amplitude quotient (AQ) is the ratio of the amplitude of vibration (see Figure 13) and the maximum declination rate of the considered signal, which is defined as the absolute value of the minimum derivative of the signal (analog to the maximum closing speed) (Titze, 2006c). This quotient can also be normalized by the cycle duration, in which case it is commonly called the normalized amplitude quotient (NAQ) (Alku et al., 2002):

$$NAQ = \frac{amplitude}{maximum\ declination\ rate \cdot t_{cycle}}. \quad \text{Eq. (6)}$$

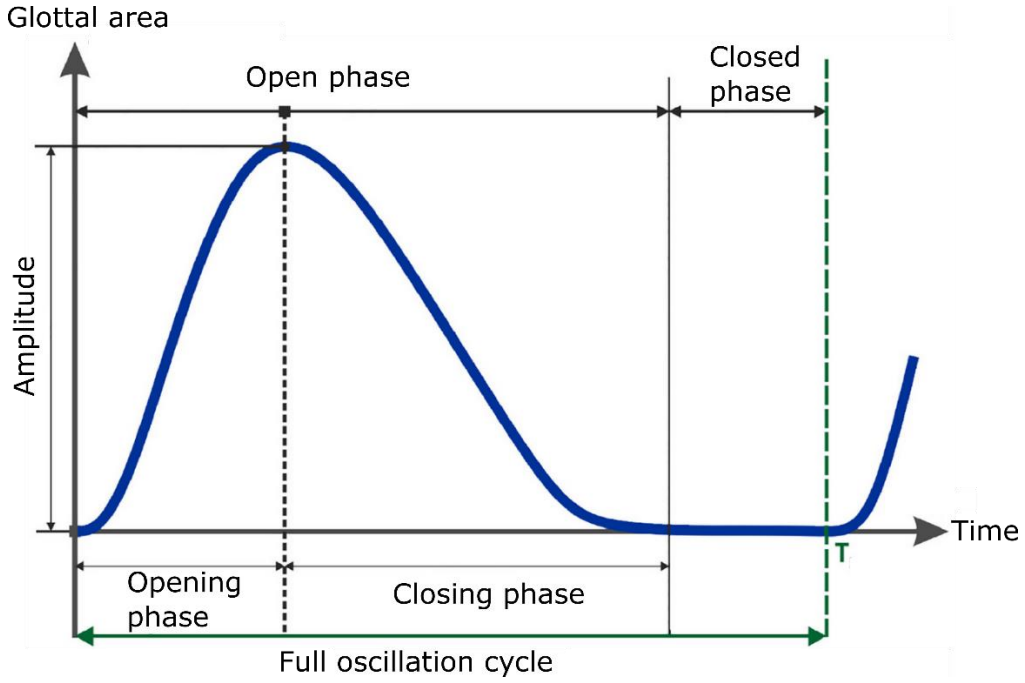


Figure 13. Description of the phases within an oscillation cycle, based on the glottal area waveform. Modified from Schlegel et al. (2019).

Finally, the kymographic contour can be used to estimate the vertical phase differences between the lower and upper margins of the VFs. The technique used to this purpose will be described in section 1.3.3.

1.2.2. Electroglottography

Electroglottography (EGG) is a non-invasive method where two electrodes measure the electrical conductance across the larynx, as represented on Figure 14a (Fabre, 1957; Baken, 1992; Herbst, 2019). Because of the differences in conductance between air and the vocal fold tissues, it is generally assumed that the EGG signal is related to the vocal fold contact area

(VFCA) (Scherer et al., 1988; Titze, 1989, 1990; Hampala et al., 2016), making it a useful tool for visualizing the VF vibrations. It is especially used to investigate the contacting (when the VFs increase contact between them) and decontacting (when the VFs decrease contact between them) instants. An idealized EGG waveform is depicted in Figure 14b: the signal commonly contains a low, flat part which corresponds to the open phase (before instant A on Figure 14b), then increases (as the lower margins of the VFs suddenly come in contact – instant A on Figure 14b) to reach a maximum value (when the contact is strongest – instant C on Figure 14b), then decreases as the VFs start decontacting, opening the glottis (instant D on Figure 14b), to finally reach a low value as the glottis is completely open (instant F on Figure 14b). The presence of a "knee" (i.e., a prominent convexity along the decontacting slope – instant E in Figure 14b) seems typical for men's voices, while women's voices appear to exhibit a smoother EGG amplitude decrease (Patel et al., 2021).

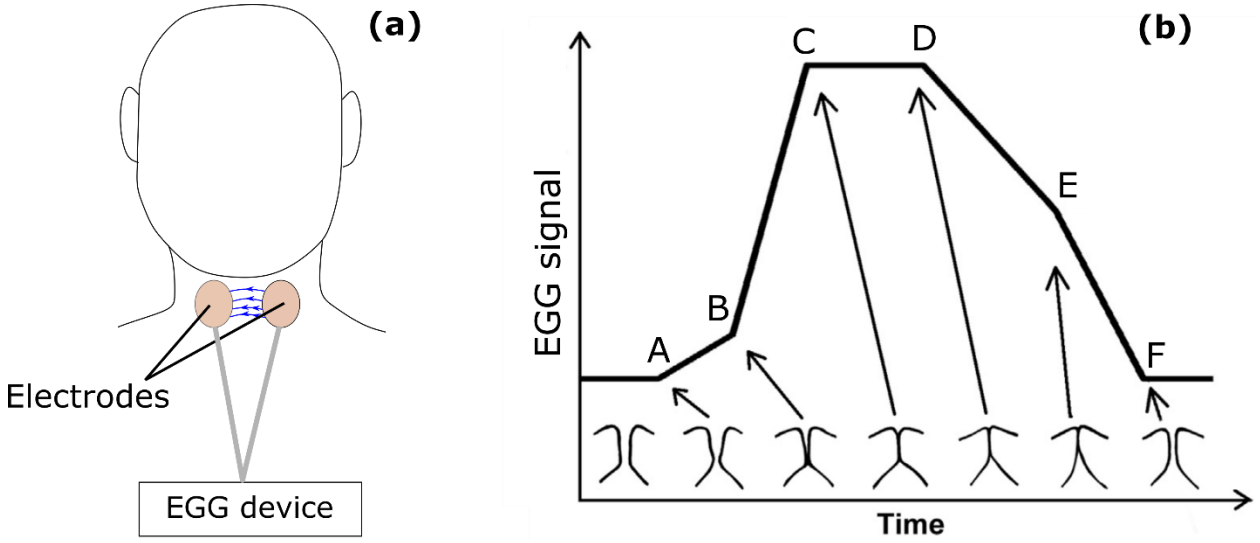


Figure 14. Description of the electroglottography (EGG) technique. (a) the two electrodes are placed across the larynx, at the position of the glottis, and connected to the EGG device; (b) idealized schematic interpretation of the EGG waveform: A – initial contact of the lower VF margin; B – initial contact of the upper VF margin; C – maximum VF contact reached; D – decontacting phase initiated by separation of the lower VF margin; E – upper margin starts to separate; F – the glottis is open, the contact area is minimal. Modified by the author from Hampala et al. (2016) and Baken (1992).

The approximate instants of contacting and decontacting can be estimated using different methods, such as considering a threshold value on the normalized EGG waveform, or using the maximum and minimum values of the first time derivative (Henrich Bernardoni et al., 2004; Herbst, 2019). Those instants are commonly used to measure the contact phase duration $t_{contact}$ and derive EGG the contact quotient CQ_{EGG}:

$$CQ_{EGG} = \frac{t_{contact}}{t_{cycle}}. \qquad \text{Eq. (7)}$$

It is sometimes expressed as the EGG open quotient $OQ_{EGG} = 1 - CQ_{EGG}$. Ternström (2019) recently proposed another method to measure the contact quotient by integrating the normalized EGG waveform, avoiding the need of determining the instants of glottis opening and closing.

The average amplitude of the EGG signal also depends on the vertical position of the electrodes on the larynx, and it will be maximum when the electrodes are placed at the level of the VFs. Slow vertical motions of the larynx can occur in speech in singing, and therefore a drop in the average amplitude could either suggest a decrease in the VF contact strength, or a vertical larynx motion, or both. To overcome this limitation of EGG, some modern EGG devices also include multiple-channel electrodes in order to detect vertical motions of the larynx and estimate its relative vertical position with regards to the electrodes (Rothenberg, 1992; Kob et al., 2009).

1.2.3. Acoustic recording and characteristics of the voice

Acoustic recordings of the voice sound can be used for perceptual evaluation (for example to understand how voice registers are perceived, as described in paper **III** of this thesis – see also section 1.4 regarding voice registers). They are also used for acoustic analysis to investigate acoustic properties of the voice sound, such as its spectral content (i.e., the energy distribution as a function of frequency), its fundamental frequency and intensity, their variations and perturbations in time, etc. This section first describes the basic acoustic features of the voice sound, and then describes parameters that can be measured through acoustic analysis.

1.2.3.1. Source-Filter theory of voice production

The basic acoustic theory of voice production is called the "source-filter" theory (Fant, 1960). This theory describes the voice production as two systems: the source (airflow modulated by the VF oscillations – see Figure 15a-b) and the filter (vocal tract resonances – see Figure 15c-d). The frequency spectrum of the source glottal flow exhibits several harmonics (peaks in the spectrum – sometimes referred to as "partials"), which are multiples of the lowest harmonic (Figure 15b). In regular voices, the frequency of the lowest harmonic normally corresponds to the fundamental frequency of oscillation f_o (Titze et al., 2015). As mentioned in section 1.1.1.2, the supraglottal tract presents acoustic resonances which act as a filter on the source sound (Figure 15d), and the resonance frequencies depend on the supraglottal tract geometry (Figure 15c). The output voice sound is the result of the linear combination of the source and the filter (Figure 15e-f), and the local maxima in the envelope of the spectrum are called formants (Titze et al., 2015).

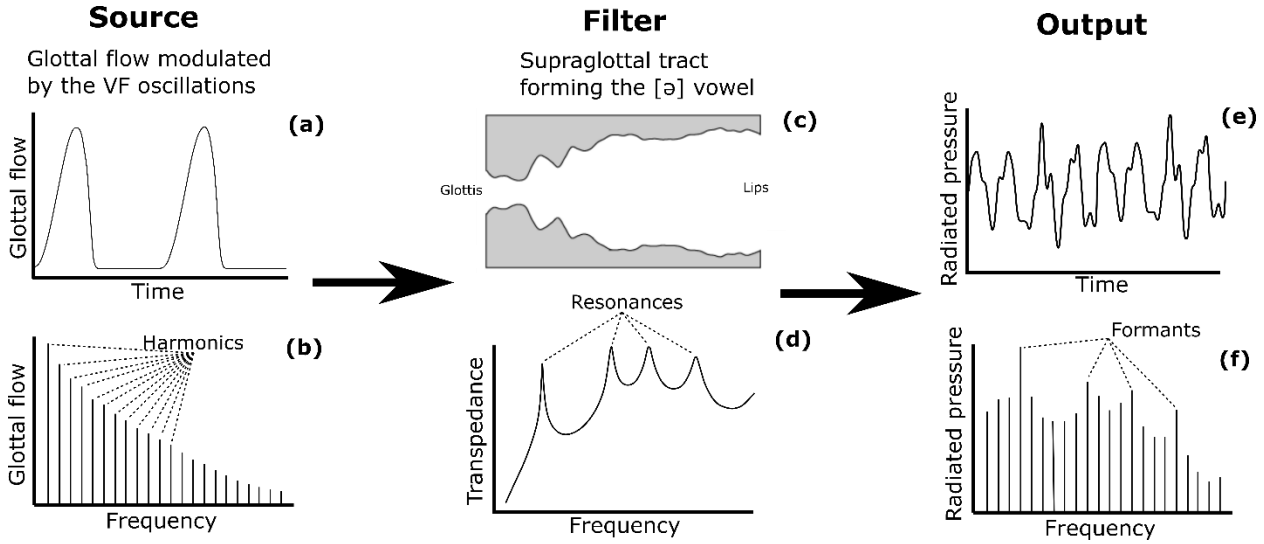


Figure 15. Description of the source-filter theory of voice production. (a) Source: glottal flow waveform generated by the VF oscillations; (b) Spectrum of the glottal flow; (c) Filter: geometry of the supraglottal tract; (d) Transfer function of the supraglottal tract corresponding to the geometry in (c); (e) Sound output: radiated pressure time signal; (f) Spectrum of the radiated pressure. Based on Wolfe et al. (2016), Fig. 2, modified by the author.

1.2.3.2. Acoustic analysis of the voice sound

This subsection describes two essential acoustic characteristics of voice: the fundamental frequency of oscillation f_o and the sound pressure level (SPL); additionally, the spectral parameters used in the papers of this thesis are described.

Fundamental frequency of oscillation f_o

The frequency of oscillation f_o, measured in the units of hertz (Hz), defines the number of oscillation cycles per second, and it can be measured from the voice sound. It is often associated (especially in singing voice) with the term "pitch", which is a perceptual term used to describe whether a sound is perceived to be "high" or "low". Within a musical context, this term is commonly used to refer to musical notes. In Western occidental music, the notes are called (in English) A, B, C, … G, and they are paired with a number. The number is related to a particular octave of the note: going from one octave

to a higher one means that f_o is doubled, and going to the lower octave means that f_o is halved. For example, the note A4 is commonly used for instrument tuning and is standardized to correspond to f_o = 440 Hz. Consequently, the notes A3 and A5 have the f_o of 220 and 880 Hz, respectively (see the webpages from Wolfe (2001) and Botros (2001) for the correspondences between all notes and their f_o value).

Sound pressure level (SPL)

Any sound, including the voice sound, is also characterized by how loud it is perceived. There exist several objective measures associated with the perceptual concept of "loudness", but the SPL is the most commonly used for the voice sound, both in clinical and research contexts. The value of the SPL L_p is measured using the following equation (Švec et al., 2018):

$$L_{\mathrm{p}} = 20 \log_{10} \frac{p}{p_0}, \qquad \text{Eq. (8)}$$

where p is the measured sound pressure and p_0 is the reference sound pressure which equals 20 μPa in air. As microphones usually deliver an electric signal which is amplified by an arbitrary gain, the values of the microphone signal are usually not inherently scaled with any reference. Therefore, to calculate the SPL, the microphone signal needs to be calibrated so that its values are scaled towards a reference unit, usually in the units of pascal (Pa). The SPL is commonly calculated as the average over a period of time, and a frequency weighting can be applied to approximate the sensitivity of the human ear. The extensive description of the SPL calculation and microphone signal calibration is beyond the scope of this thesis, but readers are referred to the tutorial by Švec et al. (2018) for such a description. The SPL has clinical and research relevance, as it has been shown to be positively correlated with the logarithm of the mean subglottal pressure (Bouhuys et al., 1968; Sundberg et al., 1999; Björklund et al., 2016; Sundberg, 2018), which is usually associated with vocal effort.

Voice sound spectral characteristics

The spectral content of the voice sound describes its "timbre", which is a perceptual notion. How a voice sound is perceived can be correlated with spectral features in the energy distribution along frequency. For example, a voice that sounds bright, rich, or strong, contains more energy in the high-frequency part of the spectrum, compared to a voice that sounds dark, mellow, or dull. Differences in timbre can be the result of adjustments at the laryngeal level (e.g. VF adduction) or changes in the supraglottal tract geometry, but for brevity only the former is described in this section. Differences resulting from laryngeal adjustments can be studied from the glottal source sound, which is obtained by the so-called inverse filtering technique, either from the microphone signal or from the oral flow signal captured using a special flow mask (Rothenberg, 1973; Baken et al., 2000; Alku, 2011). The process of inverse filtering generally involves removing formants frequencies from the voice sound to obtain the sound source signal which resembles the one depicted in Figure 15a. The sound source signal has notably been used in previous studies to measure the difference in the levels of the first and second harmonic L_{H1}-L_{H2}, measured from the inverse-filtered voice signal, to assess the spectral characteristics associated with different phonation types (Sundberg et al., 1999; Sundberg et al., 2001; Kreiman et al., 2007; Sundberg, 2022). Particularly, Sundberg (2022) noted that the values of L_{H1}-L_{H2} were lowest for pressed phonations and highest for breathy phonations, and that L_{H1}-L_{H2} decreased when the glottal flow amplitude was increasing. Zhang (2016), through a three-dimensional numerical model of the VFs, also reported lower L_{H1}-L_{H2} values when increasing the VF thickness (simulated by the contraction of the TA muscle – recall Figure 6).

To assess the more global differences in the voice timbre, researchers have measured and compared the energy in the low- and high-frequency regions of the voice sound spectrum. Particularly, the spectral balance (SB) is a measure of the difference between the sound levels within the high-frequency and low-frequency bands. The definition of the low-frequency and high-frequency bands seems to differ among studies, however this thesis uses the method Collyer et al. (2007), which defines the low-frequency band from 0 to 2000 Hz, and the high-frequency band from 2000 to 4000 Hz. The

SB can reflect the overall timbral quality of the voice sound, and it was found to be positively correlated with the SPL, indicating that increasing the SPL is related to increasing the energy in the high-frequency bands and flattening the spectral slope (Collyer et al., 2007, 2009).

1.3.Myoelastic aerodynamic (MEAD) theory of voice production

In practice, the source-filter model does not attempt to explain the mechanisms of VF self-oscillations, as well as various nonlinear dynamic phenomena of voice. For this purpose, the MEAD theory of voice production was explicitly formulated by Van den Berg (1958) and refined throughout the years by e.g., Titze (1980, 1988b, 2006b) and Švec et al. (2021). This theory opposed the so-called neuro-chronaxic theory, which assumed that the VF oscillation was caused by very fast contractions of the TA muscle (Husson, 1950).

The MEAD theory states that the VFs are driven by a combination of aerodynamic forces and of the elastic properties of the VF tissue. The aerodynamic forces are generated by the airflow created from an increase in the mean subglottal pressure, and the elastic properties of the VF tissue can be modified through the activity of laryngeal muscles. The elastic properties of the VFs play a role in the frequency of VF oscillation f_o, which is described in the following subsection.

1.3.1. Control of f_o

To understand how the properties of the VFs affect their oscillation frequency, the VFs can be modelled as strings or as simple lumped element (mass-spring-damping) systems. Notably, Titze et al. (1988) and Verdolini et al. (1994) applied the string equation, which relates the frequency of oscillation f_o to the VF length L, the VF tissue density ρ, and the VF tensile stress σ:

$$f_o = \frac{1}{2L}\sqrt{\frac{\sigma}{\rho}}. \qquad \text{Eq. (9)}$$

This equation, as an approximation, helps to understand that f_o is inversely proportional to the length of the VFs, and proportional to the square root of the stress σ.

Increasing f_o can be achieved by increasing the stress σ in the VF tissues, which can be done by two different ways:

- Activating the TA muscle, which bulges the VFs and actively tenses the muscle layer (see Figure 6).
- Activating the CT muscle, which pulls the thyroid cartilage forward and rotates it with respect to the cricoid cartilage (see Figure 3). This elongates and tenses the VFs.

The TA and CT muscles are antagonists, which means that the action of one counteracts the action of the other. In practice, varying f_o often results from a combination of the activity of both muscles. Figure 16 demonstrates the theoretical effect on f_o from possible combinations of the CT and TA muscle activities (Titze et al., 1988; Titze et al., 1989; Titze, 2000b). From Figure 16, it is clear that the action of each muscle has a different impact on the variation of f_o. In particular, two possible paths to increase f_o are shown as black arrows on Figure 16. The path which starts lower than the diagonal (straight arrows) represents a greater activity in the TA muscle than in the CT muscle. However, once the TA muscle reaches its maximum activity, no more increase in f_o can be achieved in this direction. The solution to further increase f_o is usually to abruptly disengage the TA muscle. This is assumed to create a voice break and a sudden jump to a higher f_o, as indicated by the second straight arrow on Figure 16. This kind of jump is commonly referred to as a frequency jump, or sometimes pitch jump. It is typically observed during register transitions (registers will be described in more details in section 1.4). Frequency jumps can be cultivated in certain styles such as yodeling, but singers are more commonly trained to avoid them and to execute a smooth f_o increase, without any perceivable transition. This can be performed by following the other path depicted by the curved arrow in Figure 16: the activity of the CT muscle needs to be gradually greater while gradually disengaging the TA muscle (Titze, 2000b).

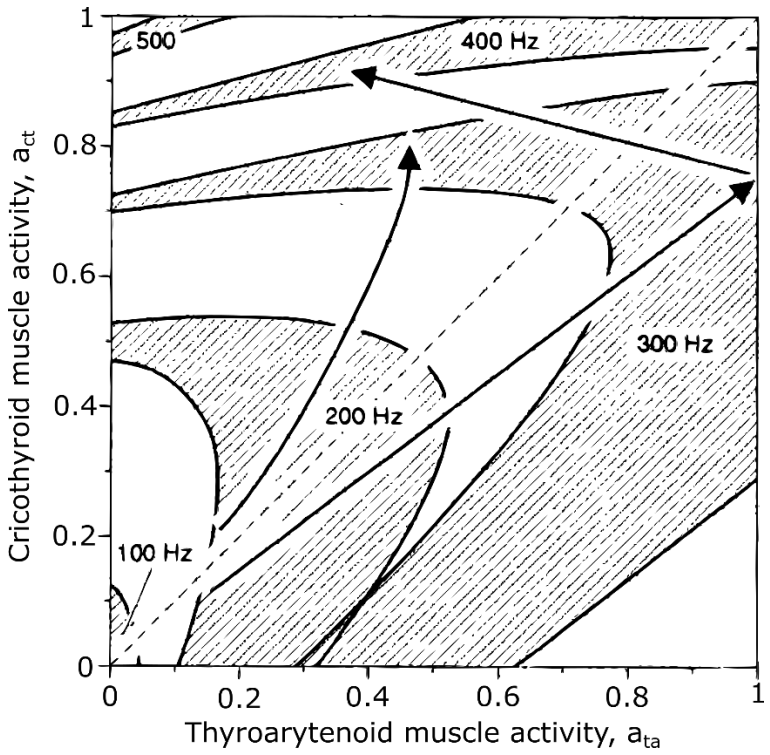


Figure 16. Muscle activation plot (MAP) indicating theoretical effects of CT muscle activation (a_{CT}) and TA activation (a_{TA}) on f_o. The two straight arrows and the curved arrow indicate different paths to increase f_o. Modified from Titze (2000b), Fig. 10.11.

1.3.2. Acoustic interactions with the supraglottal and subglottal tract

Generally, the MEAD theory suggests that the VF vibrations are influenced by the oscillating acoustic pressures below and above the VFs (Herbst et al., 2023). Titze (2008) proposed to separate the acoustic interactions between the VF vibration and the supraglottal or subglottal pressures into two groups, denoted as level 1 and level 2 interactions. Level 1 interaction occurs when the supraglottal or subglottal tract acoustic properties affect the glottal flow waveform, for example when the flow pulse is skewed from the effect of the supraglottal tract inertance (Rothenberg, 1981; Titze, 2006c). Level 2 interaction occurs when the supraglottal or subglottal tract acoustic properties affect the vibrational pattern of the VFs, the most notable example of such interaction being the occurrence of frequency jumps when f_o is in the vicinity of the first supraglottal tract formant frequency. This has been demonstrated both in vivo and in numerical simulations (Titze, 2008; Titze et al., 2008; Wade et al., 2016, 2017; Murtola et al., 2018). An example of such frequency jumps is shown in Figure 17: the lowest trace in the spectrogram shows the value of f_o and the horizontal line displays the estimated frequency of the first supraglottal formant (here denoted as F1). Small black arrows indicate sudden frequency jumps, and the arrows #1 and #3 clearly indicate that the value of f_o suddenly jumped from above to below F1 (arrow #1) and from below to above F1 (arrow #3). This suggests that the first supraglottal resonance pushed the VF vibrations away from vibrating at the frequency of the first supraglottal tract resonance.

Besides of level 2 VF interactions with the supraglottal tract, similar effects have also been hypothesized by Titze (1988a) to occur from VF interactions with the subglottal tract. It has been experimentally (with a physical model of the VFs) demonstrated by Zhang et al. (2006) that the subglottal resonances can also entrain the frequency of oscillation of the VFs. The influence of the subglottal resonances on the VF oscillations is further investigated with excised larynx experiments (to be described in section 1.5) in papers **I** and **II** in this thesis, which are related to level 1 and level 2 interactions with the subglottal tract.

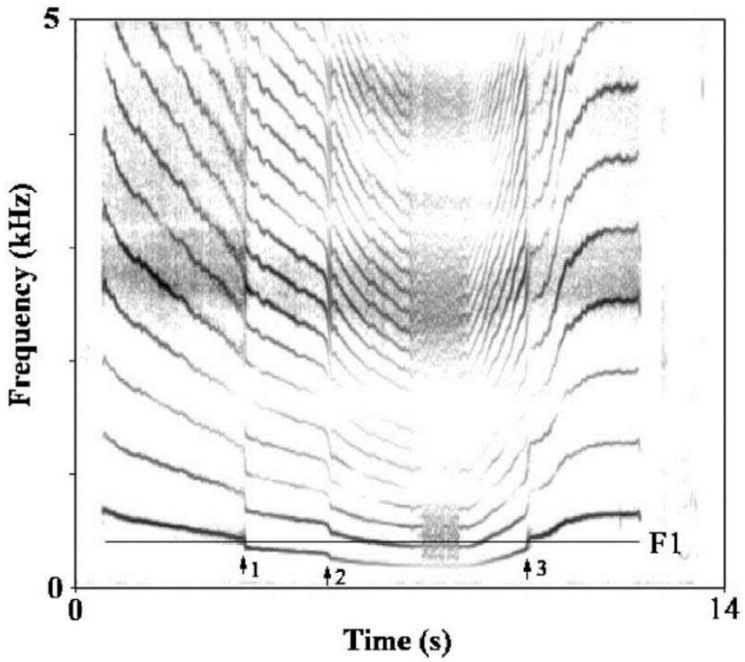


Figure 17. Spectrogram of a downward and upward glissando, showing occurrences of frequency jumps (see vertical arrows). The horizontal line indicates the estimated frequency of the first formant (F1). Reprinted from Titze et al. (2008), Fig. 5.

The inertance of the supraglottal tract has been shown to play a crucial role for level 2 interactions and consequently also for driving the VF oscillations: the inertia of the air in the supraglottal tract generates a positive supraglottal pressure during the glottal opening and a negative supraglottal pressure during glottal closing, which helps to transfer the aerodynamic energy into the VFs for self-sustained oscillations (Rothenberg, 1981; Titze, 2000d; Švec et al., 2021). This phenomenon was demonstrated already in the numerical one-mass model by Flanagan et al. (1968), which was unable to vibrate without a supraglottal tract. However, excised larynx experiments showed that VFs can exhibit self-sustained oscillations even without a supraglottal tract, which suggests that there exists another mechanism that drives the VFs and provides a sufficient energy transfer from the airflow. This mechanism is related to the nature of the VF vibration pattern, particularly with the vertical phase differences related to the mucosal waves that propagate along the VF surface. This mechanism of self-oscillation is described in the next section.

1.3.3. Vocal fold mucosal waves and vertical phase differences

For the VFs to self-sustain oscillations, the energy transfer between the airflow and the VF tissue must be sufficient to overcome the natural damping caused by the tissue viscosity. Titze (1988b) argues that, without the inertance of a supraglottal tract, the energy transfer is not sufficient to sustain oscillation if the shape of the VF medial surface does not change during the VF oscillatory phases, as depicted in Figure 18a. Instead, the shape of the VF medial surface must exhibit a change between the opening and closing phase, such that it is convergent during the opening phase and divergent during the closing phase, as depicted in Figure 18b. This pattern of vibration is generated by a vertical phase difference between the vibration of the lower margin and the upper margin of the VFs. Ishizaka et al. (1972) demonstrated this pattern of vibration with a two-mass model of the VFs, which was capable of sustaining oscillations even without a supraglottal tract. The vertical phase differences create surface waves that propagate along the medial and top surface of the VFs, which are called mucosal waves.

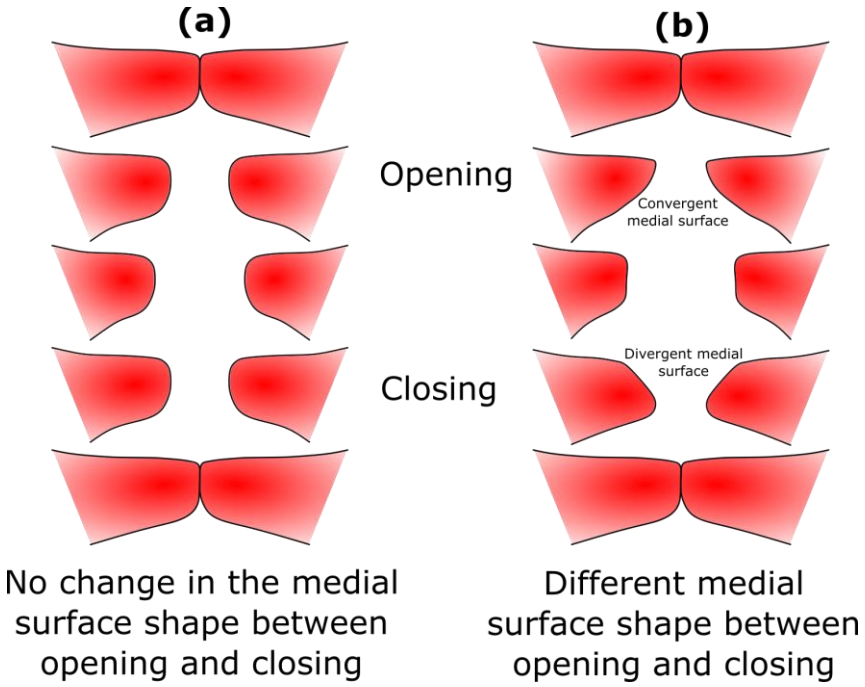


Figure 18. Oscillation cycle of the VFs. (a) Constant shape of the VF medial surface: the VFs vibrate as a whole, without any vertical phase difference; (b) the VFs vibrate with a vertical phase difference which creates a convergent shape during the opening phase and a divergent shape during the closing phase.

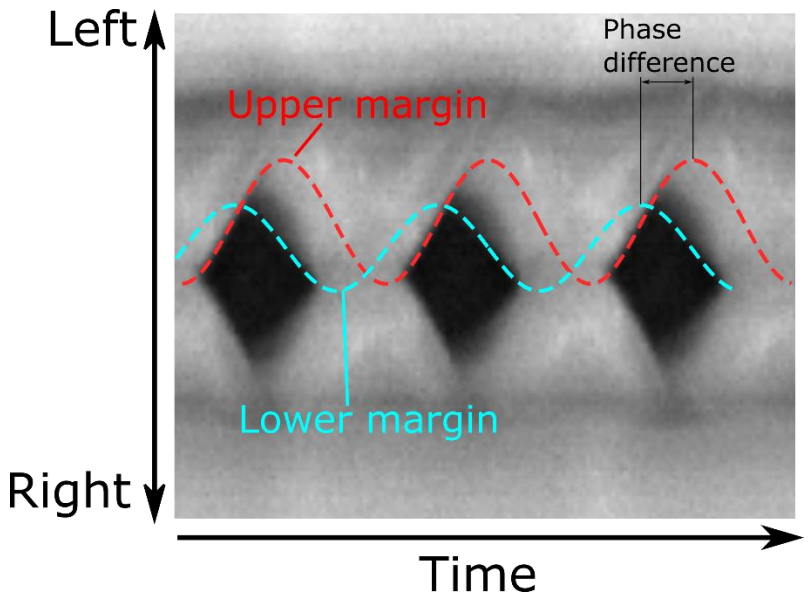


Figure 19. Kymogram of a steady phonation, showing sinusoid curves fitted to the upper (red curve) and lower (teal curve) margins for the right VF.

The mucosal wave behavior can also be qualitatively assessed by looking at the sharpness of the lateral peaks on the kymographic contours (Hiroto, 1966; Sundberg et al., 2001; Švec et al., 2007; Krausert et al., 2011; Phadke et al., 2017; Kumar et al., 2020): the sharper the lateral peaks, the higher the vertical phase difference between the lower and upper margin. Generally speaking, a higher vertical phase difference (and therefore, a lower mucosal wave velocity) indicates a high pliability of the VF mucosa, which is considered as an indicator that the mucosa is rather healthy and not damaged. The mucosal wave behavior has also been investigated in modelling studies, which will be described later in section 1.6.

1.3.4. Eigenmodes of vibration

The oscillating motion of the VFs can be decomposed into a series of natural modes of vibration, which are commonly referred to as eigenmodes, while each eigenmode is associated with a natural frequency of oscillation, called eigenfrequency (see e.g., Kinsler et al. (2000) for a general description of modes of vibration). The VFs possess an infinite number of possible eigenmodes, however it was shown that in regular phonation, two eigenmodes are often sufficient to explain a high proportion of the vibration (Berry et al., 1994; Berry & Titze, 1996; Berry, 2001).

During regular phonation (sometimes referred to as modal phonation), the vibration is periodic (with some slight degree of irregularity in the period and the amplitude). In this case, the principal eigenmodes of vibration are synchronized: their respective eigenfrequencies are very close together. This phenomenon is referred to as eigenmode entrainment by Berry (2001), and it is a nonlinear phenomenon facilitated by e.g., vocal fold collision and aerodynamic forces. A desynchronization of the eigenmodes (i.e., the eigenfrequencies are farther away from one another) causes the phonation to be no longer regular, evoking vocal instabilities (Herzel et al., 1994). This phenomenon offers another mechanism to explain the occurrences of sudden frequency jumps described in section 1.3.1: a small variation of input parameters such as the VF tension could cause a desynchronization of the eigenmodes, abruptly followed by an eigenmode entrainment to another frequency.

In the field of nonlinear dynamics, a system abruptly switching from one regime to another when input parameters are smoothly varied is called a bifurcation. Bifurcations occur within ranges of parameters where two or more dynamic regimes coexist (Berry, Herzel, et al., 1996). Notably, frequency jumps are often described as bifurcations, where the oscillations suddenly change their frequency to another.

1.4.Registers

The concept of voice registers is a controversial topic in the voice research field, and there seems to be no universal consensus on the definition and the number of registers (Mörner et al. (1963) listed more than a hundred different terms to refer to registers). The use of the term “register” was originally borrowed from organ terminology (Merkel (1863), as cited in Large (1972)): the differences in timbre between regions of notes sung on each side of the passagio (transition tones where a register change typically occurs) were analogous to different sets of organ pipes called registers, and the term was then used also for singing (Alderson, 1979).

The timbral differences have therefore been used as a way to define and describe registers. Other approaches include using proprioceptive sensations, for example the sensation of “resonance” in certain parts of the body, giving rise to the terms chest register and head register (Stark, 1999), or the differences in the vibratory pattern of the VFs (Henrich Bernardoni et al., 2003; Henrich Bernardoni, 2006; Roubeau et al., 2009). This latter approach was originally formulated by Manuel Garcia (1847), who described registers as a series of “consecutive and homogeneous tones […] produced by the same mechanical principle, and whose nature differs essentially from another series of tones equally consecutive and homogeneous produced by another mechanical principle”. It is generally recognized that there are four main registers:

- Vocal fry (or M0, pulse, strohbass): describes pulse-like VF vibrations, with a low f_o, typically below 70 Hz (Titze, 1988a).
- Chest (or M1, modal): describes the main register used for speech and the lower singing range.
- Head (or M2, falsetto, loft): describes the register used for the upper singing range.
- Whistle (or M3, flute, flageolet): describes the register used for the extreme high end of the singing range. The sound is very flutey, like a whistle (Švec et al., 2008; Garnier et al., 2012).

Other approaches exist, based on pedagogic experience, that define more than four registers (Sadolin, 2000) or a mix between two registers (Castellengo et al., 2007), but the description of those approaches is beyond the scope of this thesis.

1.4.1. Laryngeal mechanisms

Roubeau et al. (1987) were among the first to use the term “mechanism”, particularly to refer to the chest (mechanism I) and head (mechanism II) registers. Later, the term “laryngeal mechanism” (and the associated terms M0, M1, M2 and M3) was proposed by Henrich Bernardoni et al. (2003). This terminology was used to suggest that there are distinct vibratory regimes of the VFs, characterized by an observable transition between them. This transition can take the form of a frequency jump or an abrupt change in the amplitude of the EGG signal (recall section 1.2.2) or in the EGG contact quotient (Henrich Bernardoni et al., 2003; Henrich Bernardoni et al., 2004, 2005; Roubeau et al., 2009). Roubeau et al. (2009) suggest that there can be great timbral and intensity variation in the sounds produced within one laryngeal mechanism, and that even though the notions of register and of laryngeal mechanism overlap, they should be treated as different. They also suggest that each laryngeal mechanism is somewhat limited to a certain frequency range, even though two neighboring laryngeal mechanisms can have partially overlapping frequency ranges.

1.4.2. Characteristics of chest and head registers

As previously mentioned, one possible way of recognizing registers is to use one’s own kinesthetic sensations, the development of which is commonly done through singing training. Singing teachers usually provide feedback based on the sound that is produced during singing practice, considering that their perception of this sound is shaped by their own practice and pedagogic experience. Nevertheless, the characteristics of registers have been studied through various methods, which are described in this section. For brevity, this section will only deal with the investigation of chest and head registers.

The laryngeal differences between chest and head registers have been mostly revealed by early electromyographic studies which investigated the activity of the intrinsic and extrinsic laryngeal muscles during phonation (Hirano et al., 1970; Baer et al., 1976). In particular, those studies demonstrated that the TA muscle activity is high in chest register and falls when shifting from chest to head register. This was further supported by more recent studies measuring the activity of the TA and CT muscles (Kochis-Jennings et al., 2012; Kochis-Jennings et al., 2014). The theoretical framework of Titze (2000b) provides possible explanations for the need to disengage the TA muscle as f_o increases, as mentioned in section 1.3.1.

Physiologically, besides of increasing the tensile stress in the body of the VFs, the contraction of the TA muscle bulges the VFs (recall Figure 6), which allows for a higher membranous adduction (Herbst et al., 2011; Herbst & Švec, 2014). This can increase the duration and strength of the VF contact, which in turn generates an increase of the energy in the high-frequency harmonics, typically observed through a flatter spectral slope in chest register (Colton, 1972). In practice, this is demonstrated by the higher closed quotient values measured for chest register phonations (Sundberg et al., 2001; Roubeau et al., 2009; Echternach, Burk, Köberlein, Herbst, et al., 2017; Echternach, Burk, Köberlein, Selamtzis, et al., 2017). Values of L_{H1}-L_{H2} were also found to be lower for chest register phonations (Sundberg et al., 2001). In addition, the contraction of the TA muscle increases the thickness of the VFs (Hirano, 1974; Zhang, 2016), which can increase the vertical phase differences (recall section 1.3.3) between the lower and upper margins of the VFs. Thicker VFs can also increase the amplitude of vibration of the lower margin, which increases the glottal closing speed.

1.5.Excised larynx experiments

Despite its numerous benefits, in vivo investigation of voice production only offers limited possibilities due to ethical and technical constraints. To obtain deeper insights into the physiological and biomechanical aspects of voice production mechanisms, physical models of the larynx can be used, which offer better parameter control and more analysis possibilities than in vivo. Those models can be either synthetic (made from synthetic material such as silicone) or larynges excised after death. Both types of models have been widely used in research, but this thesis will only focus on the description of excised larynx experimental methods, which were utilized in papers **I** and **II**.

1.5.1. Brief history of excised larynx experimental setups

Early experiments from Ferrein (1741) with excised larynges were important to understand the most basic mechanisms of voice production, such as the fact that the vibration of the VFs was the source of the voice sound and that the tension applied to the VF varied f_o, although Ferrein believed that the VFs were producing sound in the same manner as strings which drive the air (Van den Berg, 1958). Later technological advances allowed researchers to perform more thorough measurements on excised larynges and to design experimental setups allowing a more robust control over various parameters. Notably, Van den Berg et al. (1959) used a setup with extensive measurement of the airflow rate delivered to the larynx and the pressure just below the glottis (subglottal pressure), while elongating the VFs (by applying longitudinal force on the thyroid cartilage and rotate it with respect to the cricoid cartilage – recall Figure 3) and applying lateral forces on the arytenoids to adduct the VFs (to simulate the action of the adductor laryngeal muscles – recall Figure 4).

The experimental setup used by Van den Berg et al. (1959) formed the basis of excised larynx experimental setups commonly used in later studies (Döllinger et al., 2011; Garcia Maxime et al., 2018). Its principle is rather straightforward: airflow is generated by a high-pressure container, then goes through another large container where it is heated and saturated in humidity, to finally be delivered to the larynx (see Figure 20). Van den Berg also included a way to measure the subglottal pressure and the temperature by connecting the subglottal tract to a U-tube and to a thermometer, as well as a system of weights and pulleys to control and monitor the forces applied for the elongation of the VF (longitudinal force) and for the arytenoid adduction (lateral force), as shown on Figure 20. The setup finally includes a microphone to register the radiated pressure, and a stroboscope to allow the visual observation of the VF vibrations.

Van den Berg further improved this experimental setup by adding sutures that mimic the action of some intrinsic laryngeal muscles (recall Figure 4). Those sutures are usually attached directly to the arytenoids, where the muscles normally attach. Such suture attachments are documented with remarkable details in the instructional film from Van den Berg et al. (1960), where the action of all three intrinsic laryngeal muscles described in Figure 4 (LCA, PCA and IA) is demonstrated, both independently and simultaneously. The action of the CT muscle is also mimicked by pulling a suture thread attached to the anterior notch of the thyroid cartilage (as depicted in Figure 3), which demonstrates an increase of f_o. Technological advances in electronics allowed recent experimental setups to include electric servo-motors coupled with force sensors, to more precisely control and monitor the lateral (adduction/abduction) and longitudinal (VF elongation) forces (Legou et al., 2015; Lagier et al., 2017; Lagier et al., 2020). This setup also controls all servo-motors centrally through an Arduino board connected to a computer, allowing more precision and reproducibility on the force controls, as well as their automation.

One common limitation of excised larynx experiments is the difficulty to mimic the action of the TA muscle. Montgomery implants (solid wedges inserted through the wings of the thyroid cartilage) have been used to adduct the membranous part of the VFs (Lagier et al., 2017), but the tensing and bulging effect of the TA muscle contraction cannot be simulated without innervation and electrical stimulation of the muscle. This was successfully performed by Berke et al. (2013), who demonstrated and documented the visible adduction motion resulting from the electrical stimulation of the TA muscle in a perfused human excised larynx. However, the procedure requires considerable preparation and knowledge regarding blood perfusion, making this kind of experiment difficult to set up.

Some substantial advances in our understanding of the voice production mechanisms were made through the use of hemilarynges, which are excised larynges where one VF is surgically removed, allowing the observation of the VF vibrations from the top and from the side (Jiang et al., 1993; Herbst et al., 2017). Notably, the relationship between the vocal fold contact area and the EGG signal was explored using hemilarynges (Scherer et al., 1988; Hampala et al., 2016). In addition, the modes of vibration of the VFs (recall section 1.3.4) were investigated by attaching micro-sutures to the VF medial surface and analyzing the motion of those micro-sutures through HSV (Berry et al., 2001; Döllinger et al., 2006; Veltrup et al., 2021).

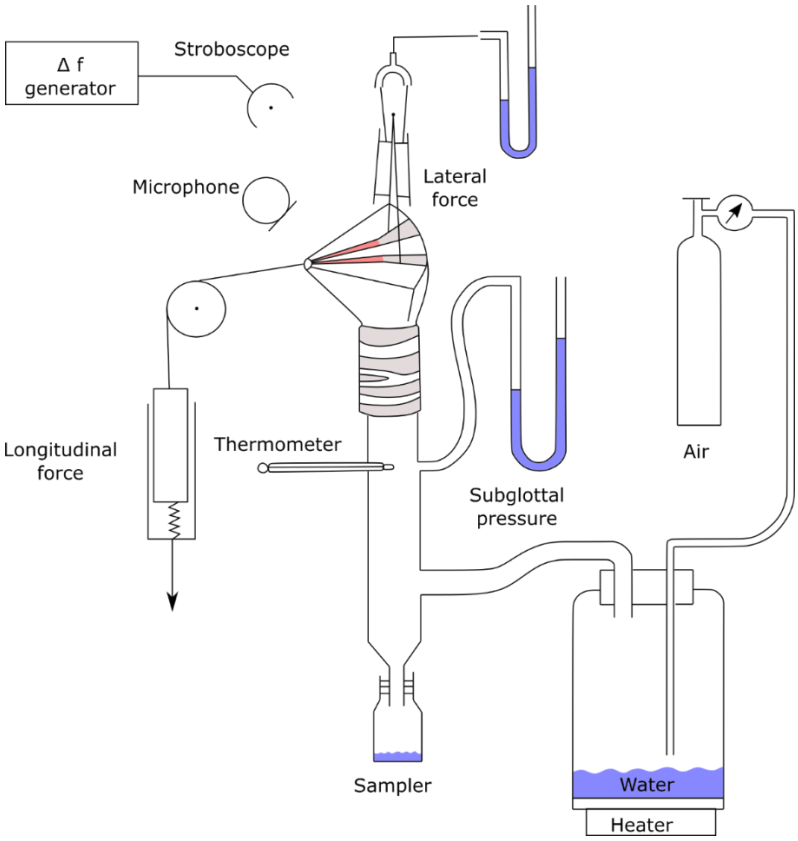


Figure 20. The setup used by Van den Berg for excised larynx experiments. Based on Van den Berg et al. (1959), Fig. 1, modified by the author.

Even though human excised larynges offer the best model for investigating human voice production and physiology, they still pose ethical concerns as they are obtained from human remains. Another limitation of excised human larynges is their age, as they are most often obtained from older individuals who died of natural cause. As age progresses, the cartilages ossify and become harder, which can make it more difficult for the preparation. To overcome those limitations, researchers have been using various mammalian larynges as replacements, such as larynges from dogs, cows, sheep, pigs, or deer. Those species seem to offer relatively similar mechanical properties of the larynx, although differences exist and should be taken into account (Alipour et al., 2013; Herbst et al., 2020). Other species have been studied, but the description of those studies is beyond the scope of this thesis. Readers can refer to the review from Garcia et al. (2018) for a more comprehensive review on excised larynx experiments.

1.5.2. Larynx storage and preparation

Larynges must ideally be excised and stored within 24 hours post-mortem (Titze, 2006a). Once larynges are harvested, they are usually flash-frozen in liquid nitrogen, which can prevent cracks, and then stored in freezers at low temperature (from -120°C to -18°C). If quick frozen in liquid nitrogen and stored properly in vacuumed bags, excised larynges can be kept for up to one month and used for experiments without substantial deterioration of the tissues altering the voice production (Chan et al., 2003). The effects of a more prolonged storage on the biomechanical properties of the VFs are unknown, however. When used for experiments, a larynx should be slowly thawed by being placed in a fridge at a temperature around 0°C overnight, and then placed in a warm water bath with saline solution at around 30°C (Baer, 1975; Durham et al., 1987; Titze, 2006a), in order to avoid any cracks or damage to the tissues caused by quick temperature variations.

Once the larynx is fully thawed, the usual preparation consists in removing superficial tissues and cartilages that are not needed for the experiment, or that can impede the view of the VFs. Most commonly, the tissues above the VFs, such as

the epiglottis, the hyoid bone, and sometimes the ventricular folds, are removed. If present, unnecessary surrounding tissues, such as the esophagus, can also be removed for facilitating the access to the laryngeal muscles and cartilages. The trachea should be kept or shortened according to the needs of the experimental setup. If needed, one VF can be removed to transform the larynx into a hemilarynx.

Once the larynx is prepared according to the needs of the experiment, it is attached to the experimental setup and the VFs should be adducted to create a sufficient airflow resistance, which will put them into oscillation. As mentioned in section 1.5.1, the arytenoid adduction can be performed by mimicking the action of some of the intrinsic laryngeal muscles (LCA and IA, for example), which is achieved through threads attached to the arytenoids and going through the tissue in the direction of the muscle. Figure 21 shows photos of an excised red deer larynx used by the author HL, with threads attached to the arytenoids to mimic the action of the LCA muscle. In this case, the threads go through the tissues in a forward and downward direction, to come out through the crico-thyroid membrane. They are also attached to a small bead to prevent the threads from losing their attachment, and to push the arytenoids towards the direction of the threads when pulled, which has the effect of rotating the arytenoids and adducting the VFs.

Another possibility for the VFs adduction is to use a system of pins or prongs that push against the arytenoids to create the adduction movement. The example of such a system is described by Titze (2006a). Figure 22 depicts the system that was used to adduct the VFs during the excised larynx experiments in papers **I** and **II** of this thesis: threaded metal prongs were screwed into a plate attached to two metal rods. Figure 22a shows the resting position of the arytenoids without the prongs exercising any force on them. Figure 22b shows the arytenoid adduction when the prongs are pressed against them: the adjustment of the prongs causes the arytenoids to come closer together and to slightly rotate.

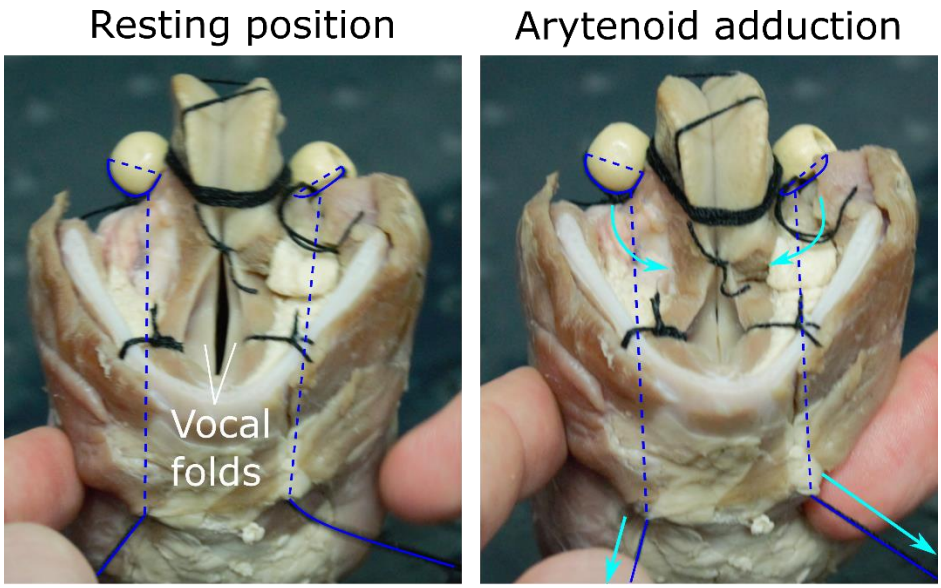


Figure 21. Excised red deer larynx with threads (in blue) attached to the arytenoids. The left panel shows the resting position of the larynx, the right panel shows the adduction process by pulling on the threads.

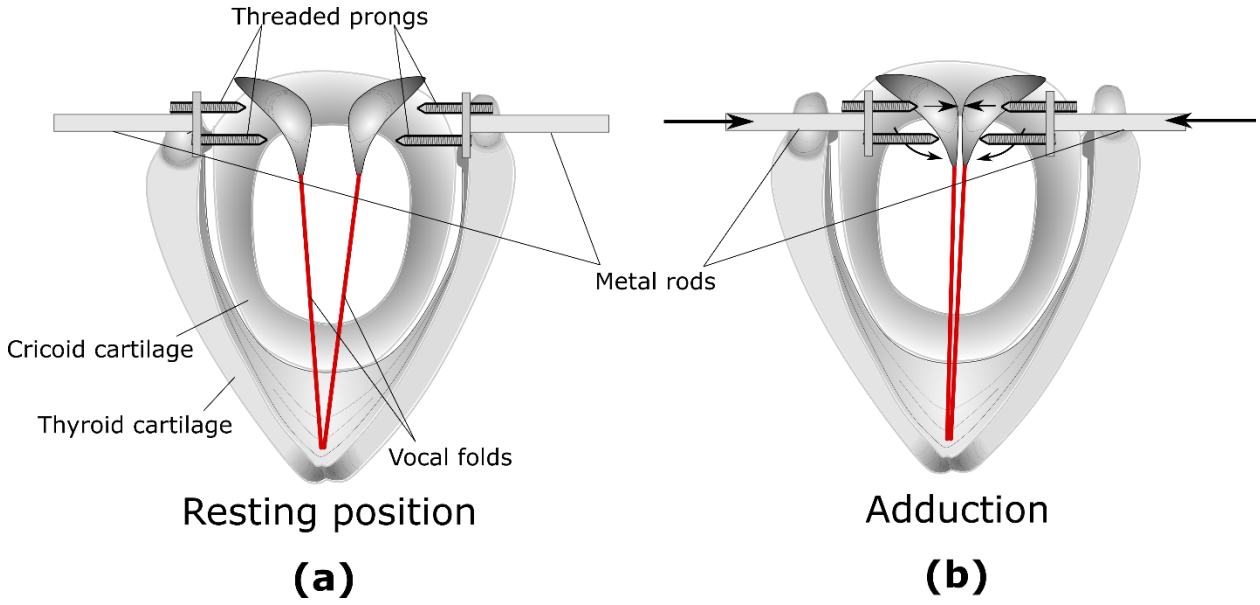


Figure 22. Description of the VF adduction using prongs. (a) resting position showing the prongs screwed into a metal rod; (b) the metal rods are pushed medially, which pushes the prongs against the arytenoids and causes them to come closer and to rotate.

1.5.3. Investigation of nonlinear dynamic phenomena in excised larynges

As previously mentioned, excised larynges offer the possibility to control, in a reproducible way, several parameters which are difficult to control in vivo. Such parameters include the airflow rate, and the VF adduction and elongation. The subglottal pressure can be indirectly controlled through the airflow rate and the flow resistance caused by the VF adduction. The degree of left-right VF asymmetry with regards to the VF adduction and elongation can also be controlled. Furthermore, the characteristics of the subglottal and supraglottal tracts, and the way they connect to the larynx, can be controlled to some extent. Therefore, excised larynges can be used to investigate nonlinear phenomena related to the inherent laryngeal characteristics and to the coupling of the VF vibrations with the supraglottal or subglottal tract.

Particularly, frequency jumps (recall sections 1.3.1 and 1.3.4) have been investigated during excised larynx experiments, by elongating the VFs (Van den Berg et al., 1960; Švec et al., 1999; Horáček et al., 2004; Tokuda et al., 2007, 2008) or by inducing a left-right asymmetry in the VF adduction and elongation (Berry, Herzel, et al., 1996). The instructional film by Van den Berg et al. (1960), particularly, clearly demonstrates frequency jumps (associated with a register transition), when the VFs are smoothly elongated by a thread pulling and rotating the thyroid cartilage. This suggests that such jumps can also occur without any acoustic interactions with the supraglottal tract. Furthermore, since the TA muscle could not be stimulated in those studies, this implies that the sudden release of the TA activity (as hypothesized by Titze (2000b)) is not necessary for frequency jumps to occur. Finally, the study from Berry, Herzel, et al. (1996) demonstrates that sufficient left-right asymmetry in the VF adduction and elongation can disrupt eigenmode entrainment (recall section 1.3.4), which stimulates eigenmode desynchronization and vocal instabilities.

1.6.Synthesis of kymograms through numerical modelling

Numerical models of the VFs can offer more control on the VF geometric and vibratory properties than physical models, and as such their use can be relevant in a research context. Specifically, numerical models can be used as a visualization tool to simulate images of the VF vibrations, such as kymograms or high-speed videos of the vibrating glottis. The development of videokymography (VKG – recall section 1.2.1.5), in the late 1990s and early 2000s, particularly stimulated the need to use numerical modelling for the generation of synthetic kymograms, to improve the interpretation of kymographic images and relate them to different types of phonation.

Notably, Horáček et al. (2009) were among the first to generate synthetic kymograms, using an aeroelastic model of the VF vibrations. This model approximates the VFs in two dimensions (2D) by a rigid body defining their shape, connected to a two-degree-of-freedom (translational and rotational) system of masses, springs and dampers (Horáček et al., 2002). Here, the VFs are driven by aerodynamic forces generated by the airflow, and kymograms were generated by placing a virtual camera above the VFs and using an illumination method to measure the quantity of light reflected by the VF surface, based on the local surface slope. Three examples of synthetic kymograms generated by this model are displayed on Figure 23. This model is able to generate simple kymograms where both the upper and lower margins of the VFs are visible (Figure 23a), however since the VFs are approximated by a rigid body, the mucosal wave extent (the propagation of the mucosal wave across the top surface of the VFs) is not visible and the kymograms are visually quite distinct from those obtained from real VFs in vivo. Nevertheless, this model allows to investigate the appearance of kymograms when varying the subglottal pressure (Figure 23b) and the frequency of oscillation f_o (Figure 23c).

Later, Švancara et al. (2014) performed a similar study with a 2D finite element model, where the VF shape is based after the M5 model defined by Scherer et al. (2001). The VFs are discretized using a finite element mesh, and set to vibration by fluid-structure interaction with the airflow. Kymograms are generated using the same illumination method used in Horáček et al. (2009). As the VFs are deformable in this model, the mucosal wave extent is visible on the generated synthetic kymograms, an example of which is displayed on Figure 24. This kymogram has a more realistic appearance than the kymogram generated by Horáček et al. (2009) but the vibratory features are still quite distinct from those visible in the kymograms of real VFs.

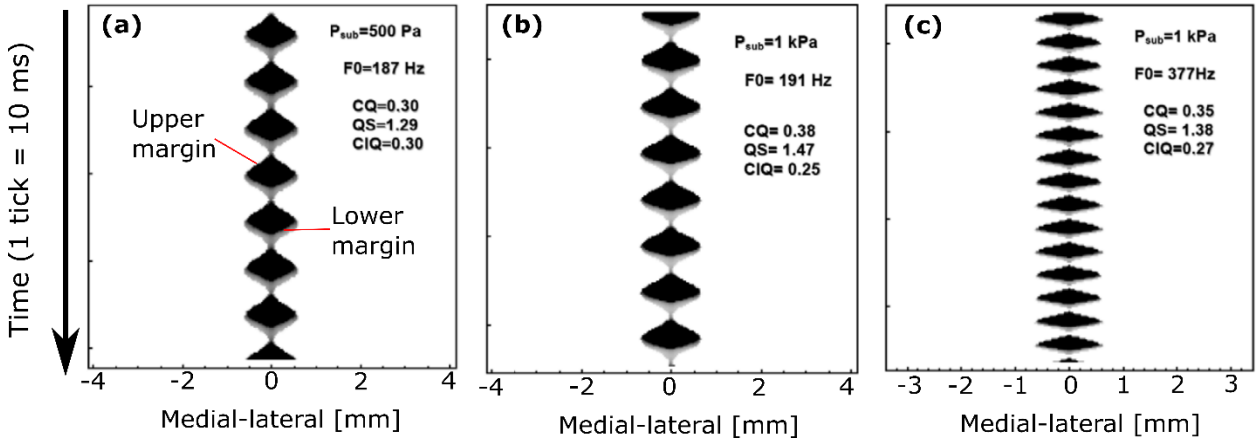


Figure 23. Example kymograms generated by an aeroelastic model of the VFs. The contributions to the kymogram from the upper and lower margins are indicated with red lines. The values of the closed quotient (CQ), the speed quotient (QS), and the closing quotient (ClQ), are also indicated. (a) Kymogram generated with a subglottal pressure of 500 Pa, with the frequency of oscillation of 187 Hz; (b) The subglottal pressure is increased to 1 kPa, which has the effect of increasing the vibration amplitude and f_o; (c) The subglottal pressure is kept at 1 kPa, but f_o is increased to 377 Hz by increasing the VF stiffness. Modified by the author from Horáček et al. (2009).

These two models involve solving complex equations of fluid-structure interactions, which are computationally expensive, especially for finite element models containing a high number of points for calculation. To address this potential limitation, kinematic VF models have been designed for simulating kymograms. Here, the VFs are set in vibration by applying kinematic rules instead of driving them with aerodynamic forces. Besides of computational simplicity, this also has the benefit of providing more direct control on the pattern of vibration. Notably, Schoentgen et al. (2015) used the phase-delayed overlapping sinusoids (PDOS) model, which applies a medial-lateral sinusoidal motion to the VF surface, where the glottal exit (upper margin) is phase delayed with respect to the glottal entry (lower margin) (Titze, 2006b). Schoentgen et al. (2015) demonstrated that the model is capable of generating synthetic kymograms exhibiting left-right asymmetry, which have a more realistic appearance since the VFs vibrate with some degree of asymmetry in vivo. However, this model cannot display the mucosal wave extent, as the motion of the VFs is only horizontal.

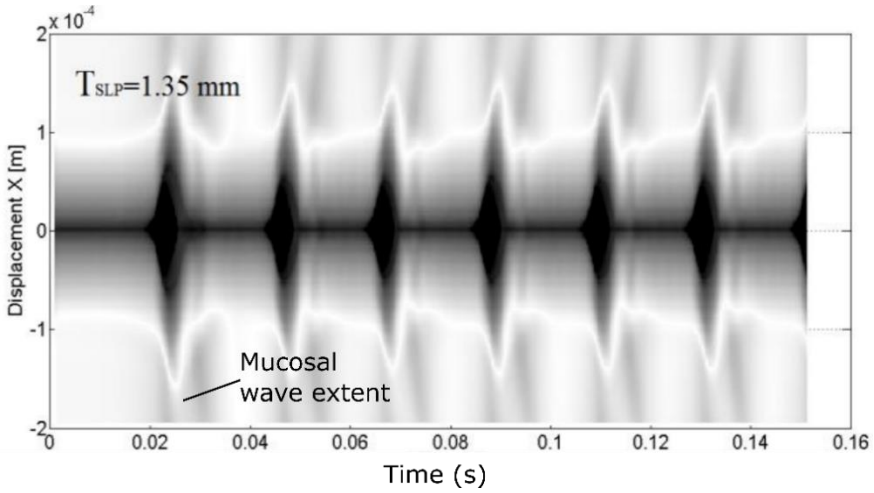


Figure 24. Kymogram generated from a finite element model of the VF vibrations. The thickness of the lamina propria T_{SLP} is indicated here. Modified by the author from Švancara et al. (2014).

The recent mucosal wave model developed by Kumar et al. (2019) addresses this issue by applying circular (lateral-medial and vertical) motion to the VFs. This model discretizes the VF surface into equidistant points following the M5 shape, which can be fully defined by a few geometric parameters. Each point is then set to motion following a circular trajectory, with a constant phase delay between all successive points. The definition of the model using only a few geometric (glottal width, convergence angle of the medial surface) and kinematic (amplitudes and frequencies of vibration, phase difference) parameters allows for a high flexibility in the appearance of the generated kymograms, which can simulate various types of phonations and offer the closest resemblance to kymograms of real VFs to date, also with regards to mucosal waves travelling over the top VF surface. This is demonstrated in Figure 25, which displays kymograms corresponding to a soft (Figure 25a), pressed (Figure 25b), breathy (Figure 25c) and head (Figure 25d) voice. This model uses a more complex illumination method which also considers the vertical distance from the camera.

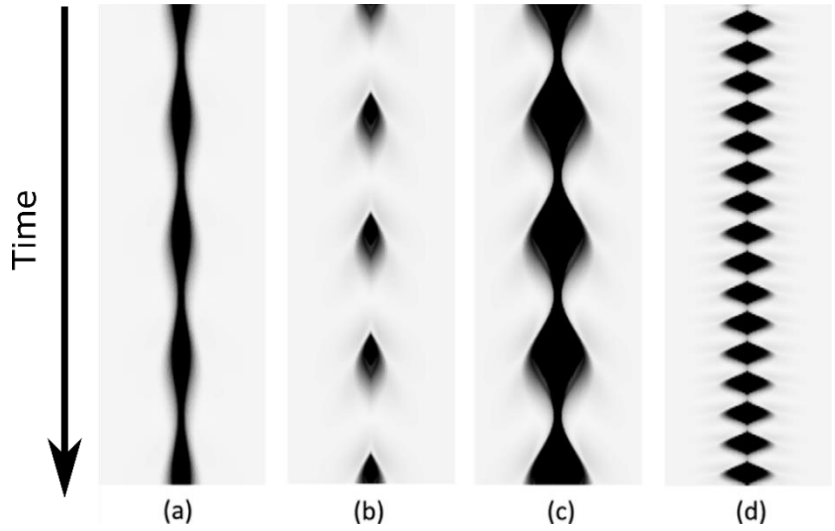


Figure 25. Synthetic kymograms generated by the mucosal wave model, showing different simulated types of phonation: (a) soft, (b) pressed, (c) breathy, and (d) head register. Modified by the author from Kumar et al. (2019).

2. ORIGINAL WORK BY THE AUTHOR

2.1. Aims of this thesis

The aims of this thesis are:

- to test and document a newly developed anechoic subglottal tract, and to investigate the behavior of the VF vibrations using excised larynges in fully anechoic conditions and compare it with subglottally resonant conditions. In particular, the aim is to document the differences in the subglottal and radiated pressure waveforms. This is addressed in paper **I**.
- to investigate the differences in the occurrences of frequency jumps in human excised larynges, between anechoic and subglottally resonant conditions. This is addressed in paper **II**.
- to investigate the laryngeal differences in vivo between chest and head registers throughout a wide singing range, and to find relevant glottal parameters that can discriminate between the two registers throughout this entire range. This is addressed in paper **III**.
- to help to extend a previously developed two-dimensional numerical kinematic model of the VF vibrations to three dimensions, so that it can generate not only kymograms but also synthetic videos of the entire glottis, which appear similar to laryngoscopic videos obtained in vivo during laryngeal HSV. Particularly, the aim is to allow the visualization of realistic vibrations, containing anterior-posterior and left-right differences. This is addressed in paper **IV**.

This section gives a summary of the work by the author on each paper in this thesis, and gives an overview of their most significant results. Readers are referred to the papers themselves for more detailed descriptions.

*2.2. Paper **I**: Development of an anechoic subglottal tract, and comparison of subglottal pressure waveforms in anechoic and subglottally resonant conditions*

2.2.1. Objectives

The glottal flow and the VF vibrations have been shown to be influenced by the acoustic resonances of both the supraglottal (Titze, 2008; Titze et al., 2008; Tokuda et al., 2010; Zañartu et al., 2011; Wade et al., 2017) and subglottal (Zhang et al., 2006) tracts. The study the inherent vibratory properties of the VFs and of the glottal voice source, it is necessary to remove all acoustic interactions with the supraglottal or subglottal tract. This study proposes to address this need by using excised larynges (which are free of a supraglottal tract) and a newly developed anechoic subglottal tract, which is described here. The influence of the subglottal resonances is further explored by comparing the subglottal pressures and properties of the VF vibrations in anechoic and resonant conditions. Specifically, the subglottal pressure waveform is investigated. In addition, the influence of the subglottal resonances on the vibratory pattern of the VFs (referred to as level 2 interactions by Titze (2008)) is assessed by measuring the EGG signal and the fundamental frequency of oscillation f_o.

2.2.2. Design of the anechoic subglottal tract

The anechoic subglottal tract was constructed in the Voice Research Laboratory at the Palacký University in Olomouc by the former doctoral student V. Hampala under the supervision of J. G. Švec. The tract consists of a 3.30 m plastic tube terminated by a wedge of polyurethane foam to absorb the sound waves propagating through it, which prevents their reflection and therefore suppresses the acoustic resonances. Detailed description, analysis of the properties and behavior of the tract was carried on by the author.

To verify that the anechoic subglottal tract was suppressing the acoustic resonances, its acoustic response was first measured and compared with the acoustic response of the resonant subglottal tract previously developed by Hampala et

al. (2013). It was found that the anechoic subglottal tract is successfully suppressing the acoustic resonances, as shown on Figure 26: the resonant subglottal tract exhibits visible resonances and antiresonances, whereas the anechoic subglottal tract is practically free of these.

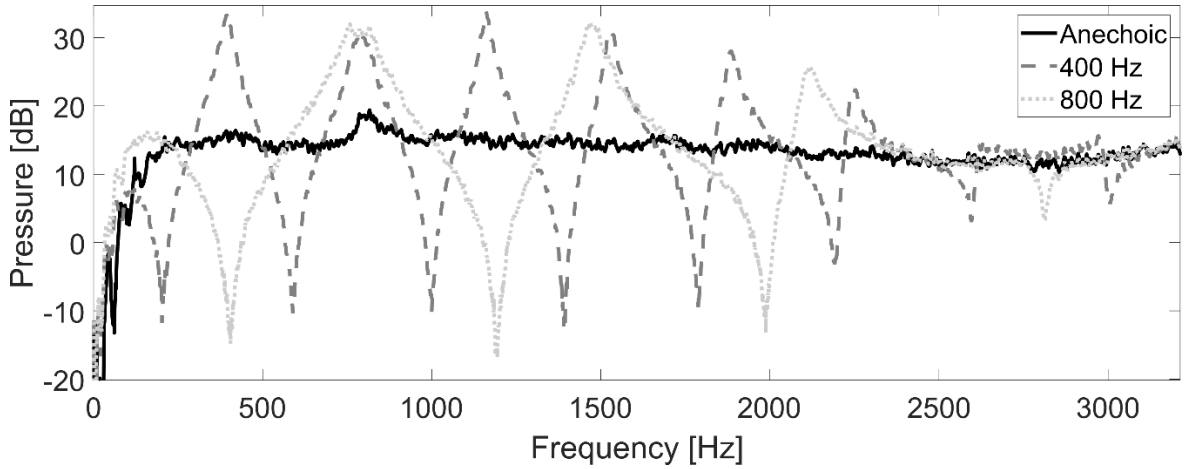


Figure 26. Frequency responses of the anechoic subglottal tract and of the resonant subglottal tract with the piston set to two different positions corresponding to fR1 = 400 and 800 Hz.

2.2.3. Excised larynx experiments

Red deer excised larynges were then explored during two experiments, using the experimental setup depicted on Figure 27. A pressure sensor was inserted through the dorsal ridge of the cricoid cartilage to measure the mean and oscillating subglottal pressure, and a microphone was placed approximately 10 cm away from the glottis to register the radiated sound. EGG electrodes were also placed on each side of the thyroid cartilage, at the level of the VFs, and connected to the Glottal Enterprises EG2-PC device to monitor the relative VF contact area. For both experiments, the value of f_o was measured using the SWIPE' algorithm developed by Camacho et al. (2008).

The first experiment investigated steady phonations (one larynx), in both anechoic and resonant conditions. In resonant conditions, the subglottal tract was set to six different resonance frequencies: f_{R1} = 330, 400, 500, 600, 700 and 800 Hz. The flow was increased to a constant mean value of about 400 mL.s^{-1}, until the mean subglottal pressure reached a saturation value, and then the flow was decreased back to 0 mL.s^{-1}. This experiment investigated the differences in the subglottal pressure waveforms between anechoic and subglottally resonant conditions.

The second experiment investigated flow sweeps (three larynges), again in both anechoic and resonant conditions. For this experiment, in the subglottally resonant conditions the resonance frequency was only set to f_{R1} = 500 Hz, to approximate the value of the subglottal resonance frequency in vivo (see section 1.1.1.3). The flow was increased slowly to the maximum value (550 mL.s^{-1}) and then slowly decreased back to 0 mL.s^{-1}. In this experiment, the phonation threshold pressure, and the SPL of the subglottal and radiated sounds, were also measured and compared between anechoic and subglottally resonant conditions – for details see the paper **I**.

Remarkable differences were visible in the subglottal pressure waveforms between anechoic and subglottally resonant conditions, shown in Figure 28. As expected, in the anechoic conditions, the subglottal pressure waveform was found to be similar to the inverted theoretical glottal flow (Sundberg, 2018), as shown on Figure 28a: it was decreasing during the opening phase, increasing during the closing phase, and was rather flat with little fluctuations during the closed phase (recall section 1.2.3.1). On the other hand, in the resonant conditions, the subglottal pressure waveforms showed strong fluctuations during the closed phase, and the frequency of those fluctuations increased as the subglottal resonance frequency increased, as demonstrated for f_{R1} = 330 Hz (Figure 28b) and for f_{R1} = 500 Hz (Figure 28c). In addition, the EGG

waveforms also exhibited, especially for f_{R1} = 330 Hz (Figure 28b), strong fluctuations in the closed phase, which appear to be synchronized with the fluctuations in the subglottal pressure waveform.

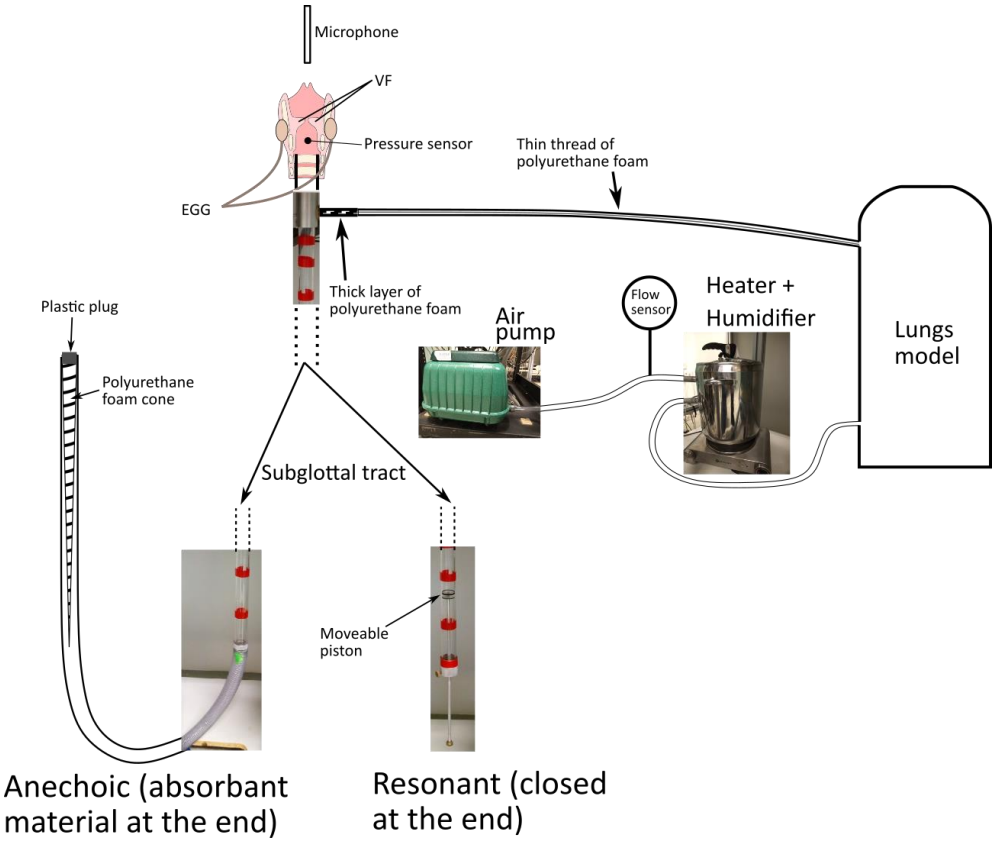


Figure 27. Excised larynx experimental setup.

In both experiments, the values of f_o were consistently lower in the subglottally resonant conditions than in the anechoic conditions. The values measured during the second experiment are plotted against the mean subglottal pressure, for larynx #1 (only the values for the first and last repetitions are shown for brevity), in Figure 29.

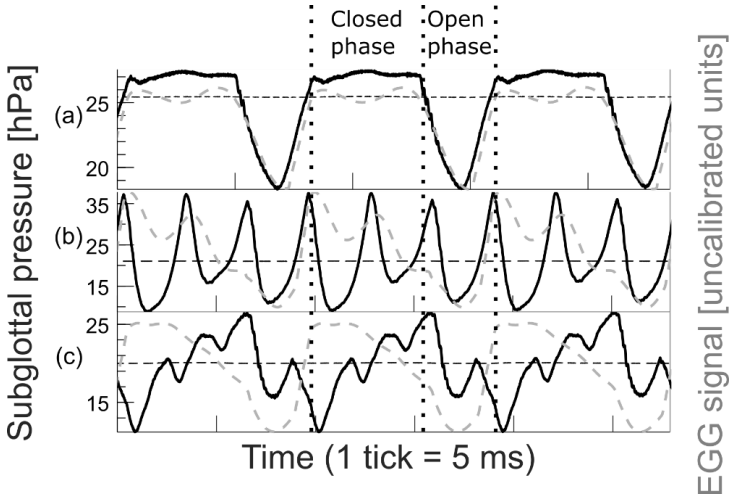


Figure 28. Subglottal pressure (solid black lines) waveforms registered with an anechoic subglottal tract and with resonant subglottal tracts set to different resonance conditions: (a) anechoic subglottal tract, (b) f_{R1} = 330 Hz, (c) f_{R1} = 500 Hz. The horizontal dashed black lines indicate the mean subglottal pressure in each case. The waveforms are individually scaled in time (3 cycles are shown). The instants of closure and opening were approximately synchronized using the corresponding EGG signal waveforms (dashed gray lines). Finally, the closed and open phases are indicated by vertical dotted black lines.

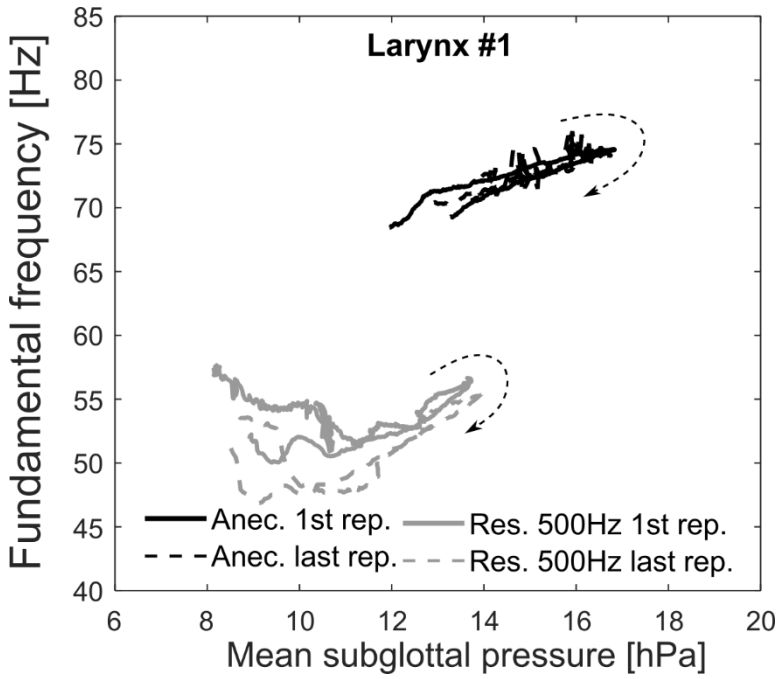


Figure 29. Measurement of the fundamental frequency of oscillation f_o made during the flow sweep experiments, for larynx #1. The dashed black arrows indicate the evolution of the measured values with time. Notice the different f_o values in anechoic and resonant conditions.

2.2.4. Discussion

The resemblance of the subglottal pressure waveform to the theoretical glottal flow waveform in anechoic conditions (recall Figure 15a) goes in line with the theoretical relationship between the flow and the pressure (described in the Appendix of paper **I**) in the absence of acoustic resonances. The flat subglottal pressure waveform during the closed phase of glottis provides a direct visual evidence of the absence of subglottal resonances and proper functioning of the anechoic subglottal tract. On the other hand, the subglottal pressure fluctuations observed during the closed phase in the resonant cases indicate the presence of subglottal resonances, as reported previously by measurements in vivo (Miller et al., 1985; Schutte et al., 1988; Švec et al., 2021).

The fluctuations in the EGG signal and the lower values of f_o measured in the resonant conditions provide direct evidence of level 2 interactions due to the presence of subglottal resonances. This corroborates the findings from Zhang et al. (2006), who also reported a strong influence of subglottal resonances on the VF oscillatory frequency f_o.

2.2.5. Conclusion

The newly developed anechoic subglottal tract successfully suppresses its acoustic resonances, removing their influence on the VF vibrations. In anechoic conditions, the subglottal pressure waveform resembles the inverted theoretical source flow waveform (recall Figure 15a), which allows to study the source signal without the need for inverse filtering. The presence of subglottal resonances consistently lowers the fundamental frequency of oscillation of the VFs and generates strong fluctuations in the EGG signal during the closed phase, which provides evidence of level 2 interactions with the subglottal pressure oscillations. The developed anechoic subglottal tract can be used to investigate the inherent vibratory properties of the voice source and the impact of structure-acoustic interactions with subglottal resonances on those properties.

*2.3.Paper **II**: Frequency jumps in anechoic conditions*

2.3.1.Objectives

Frequency jumps are a common vocal phenomenon which occurs when young prepubescent individuals go through puberty, or when singers switch from one register to another. Frequency jumps have been explained through several hypotheses: sudden release of the TA muscle (recall section 1.3.1), interactions with the supraglottal resonances (recall section 1.3.2) and with the subglottal resonances (Titze, 2000b), or nonlinear phenomenon caused by VF eigenmode desynchronization (recall section 1.3.4). However, the exact mechanism responsible for frequency jumps is still unknown. Past studies have demonstrated frequency jump events in excised larynges without supraglottal tract, which indicates that supraglottal resonances are not needed. However, as mentioned in section 2.2, subglottal resonances can have an influence on the VF vibrations, and as such could be responsible for the occurrence of frequency jumps in excised larynges. Consequently, the question of whether frequency jumps can also occur in fully anechoic conditions remains unanswered. This study proposes to answer this question by exploring frequency jumps in human excised larynges in anechoic conditions, using the anechoic subglottal tract described in paper **I** (see section 2.2 for a brief summary) and no supraglottal tract. The influence of the subglottal resonances on frequency jumps is also investigated by using a resonant subglottal tract.

2.3.2.Methods

Five human excised larynges are used with the experimental setup described in paper **I** (see Figure 27). The frequency of oscillation f_o is increased and decreased by smoothly elongating and shortening the VFs, which is done by manually pulling the anterior notch of the thyroid cartilage (see Figure 30). Elongation/shortening sweeps are performed in anechoic condition and in subglottally resonant conditions with the first subglottal resonance frequency set to f_{R1} = 500 Hz (i.e. similarly as observed in the *in vivo* conditions), and repeated to assess reproducibility

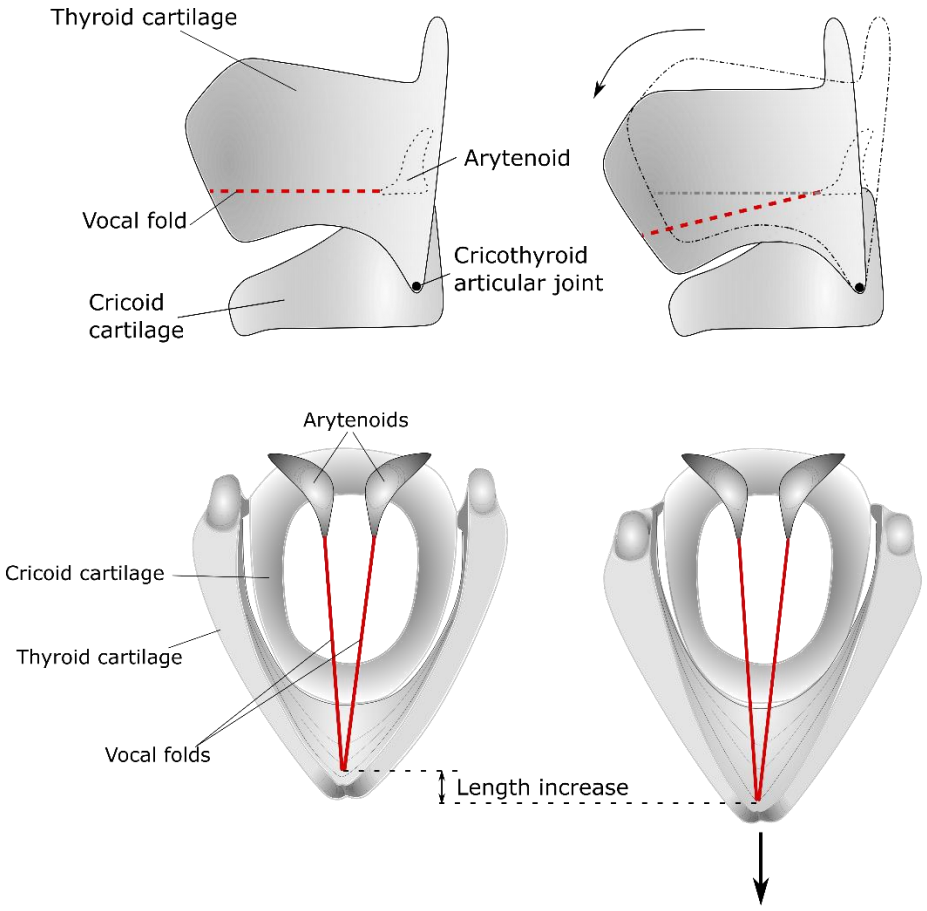


Figure 30. Description of the VF elongation process by pulling on the anterior notch of the thyroid cartilage.

2.3.3.Results

When smoothly elongating the VFs, consistent jumps were observed in both anechoic and subglottally resonant conditions, for two larynges. Figure 31 shows the values of f_o right before and after each jump, for larynx #1 and larynx #2, in both anechoic and resonant conditions. The presence of subglottal resonances did not stimulate more numerous jumps than in anechoic conditions, however the resonant subglottal tract slightly altered the starting and terminating frequencies of the jumps, especially during the upward jumps (Figure 31a,c).

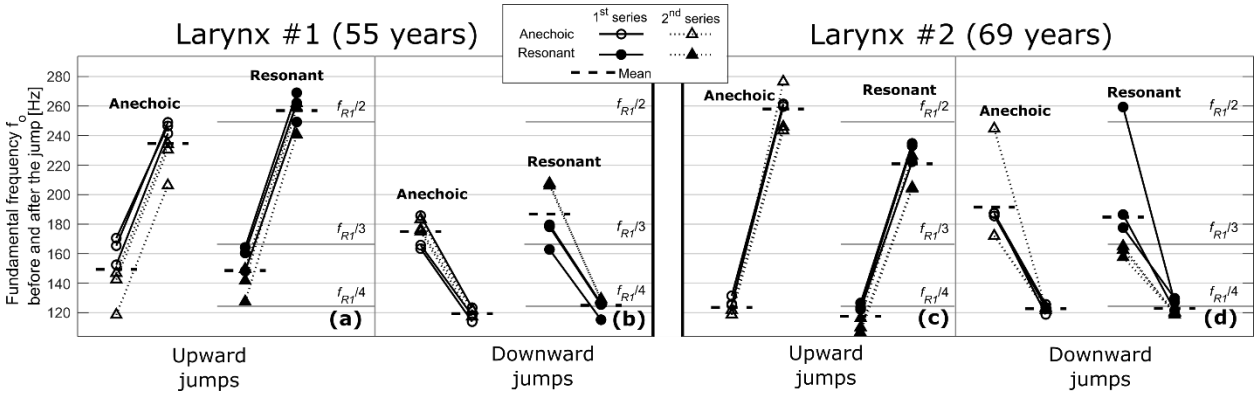


Figure 31. Fundamental frequency (f_o) before and after each frequency jump for larynx #1 and #2. The circles correspond to the first, and the triangles to the second series of frequency sweeps. For the resonant cases, we also indicate the integer divisions of the first subglottal resonance frequency ($f_{R1} \approx 500$ Hz), $f_{R1}/2$, $f_{R1}/3$, and $f_{R1}/4$. (a) larynx #1, upward jumps; (b) larynx #1, downward jumps; (c) larynx #2, upward jumps; (d) larynx #2, downward jumps.

2.3.4.Discussion and conclusion

This study presents the first clear evidence that frequency jumps occur in excised human larynges also in fully anechoic conditions, suggesting that acoustic interactions with the supraglottal and subglottal tracts are not necessary for those jumps to occur. Furthermore, the presence of acoustic resonances in the subglottal tract did not stimulate more numerous jumps compared to anechoic conditions, which confirms that those jumps were primarily caused by inherent VF properties rather than by acoustic interactions with the subglottal tract. Nevertheless, the presence of acoustic resonances in the subglottal tract had an impact on the terminating frequencies of the upward jumps, which suggests that the subglottal resonances also have an influence on the frequency jumps, although this influence appears secondary. As such, this study provides important insights on the underlying mechanisms behind frequency jumps.

*2.4. Paper **III**: Laryngeal adjustment differences between chest and head registers*

2.4.1.Objectives

As mentioned in section 1.4, the exact underlying mechanisms behind the production of voice registers are not fully understood. Frequency jumps described in paper **II** (see section 2.3 for a brief summary) are often associated with a transition between two registers, even though singers are able to operate a smooth pitch glide without any perceivable transition event, by precisely controlling the adjustments of the laryngeal muscles (recall section 1.3.1). These adjustments, and particularly the activity of the TA muscle, have been hypothesized to be essential in producing different registers, specifically chest and head registers. Previous studies have investigated the differences in laryngeal adjustments between chest and head registers (recall section 1.4.2), but only over a limited pitch range where both registers can be produced. However, the pitch limitation of chest register to low pitches and of head register to high pitches could be challenged. There is even evidence that both registers can be produced over a wider frequency range, for example as demonstrated by the Australian singer Mal Webb, who performs register transitions while singing the same note, in his YouTube video “Sideways yodeling” (Webb, 2008).

Overall, a way to reliably recognize chest and head registers, and differentiate one from the other independently from pitch, is still lacking, particularly regarding the laryngeal adjustments employed by trained singers to produce each register.

Inspired by the Mal Webb idea, the American singer Lisa Popeil came to the Voice Research Lab at the Palacký University in Olomouc with the idea of separating registers from pitch: she claimed that she can produce both the chest and head registers throughout her entire singing range. This study investigates this claim in detail, with the aim of finding relevant laryngeal and glottal parameters that can allow the discrimination between chest and head registers, regardless of pitch. The investigation is based on several hypotheses based on previous studies reporting more TA muscle activity and a richer harmonic structure of the sound in chest register. Those hypotheses are described in detail in paper **III** in this thesis but can be summarized as follows: in chest register, the VFs will exhibit a longer and more pronounced contact, larger phase differences between their lower and upper margins, a higher maximum closing speed, and there will be more high-frequency energy in the radiated sound spectrum.

In this study, the terms “chest-like” and “head-like” are used to refer to the singer’s intended registers, to avoid confusion with the pitch-associated “chest” and “head”. The terms “chest” and “head” are only used here to refer to the registers described previously in the literature.

2.4.2.Methods

The professional singer (co-author LP) performed short sustained phonations alternating chest-like and head-like registers with a short pause, on every pitch from C3 (131 Hz) to C6 (1047 Hz), except for Ab3 (208 Hz) which was omitted by mistake. The VF vibrations are captured by a high-speed camera connected to a 90° endoscope, at the rate of 7200 frames per second (FPS) for the pitches from C3 to Db5 (554 Hz), and 13600 FPS for the pitches from D5 (587 Hz) to C6. The VF contact and the radiated sound are also registered through EGG electrodes placed on the larynx and through a microphone placed on top of the camera at the distance of 22 cm from the mouth. The microphone sound is used for blind listening tests performed by the three authors of the study (“insiders” group) and by six other participants who were not familiar with the study design (“outsiders” group). The listening tests were performed to determine whether the intended registers could be perceptually distinguished. The experimental setup is shown in Figure 32.

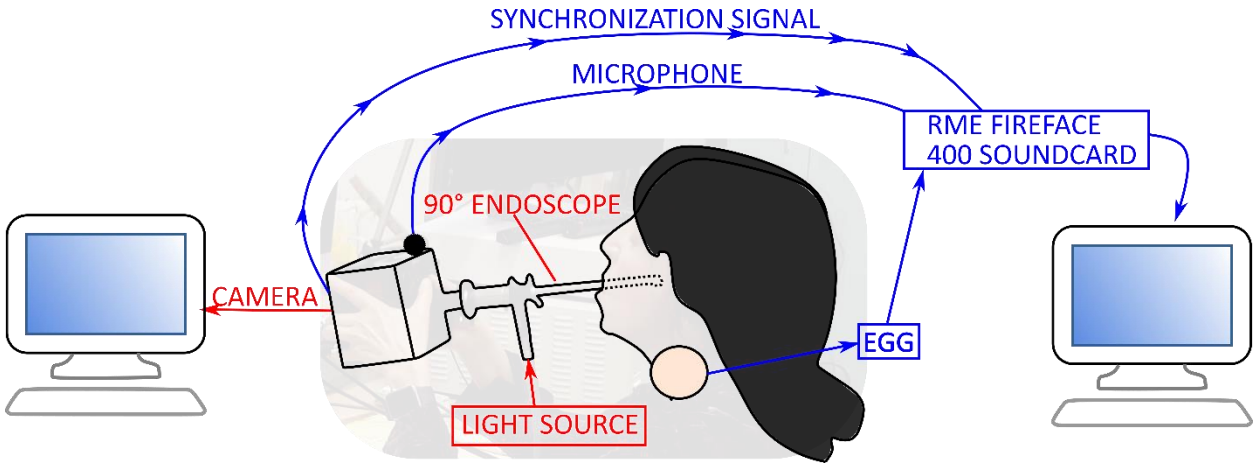


Figure 32. Experimental setup.

The following parameters are calculated from the phonovibrograms (PVG), the glottal area waveform (GAW) and the kymographic contour of the vibrating VFs:

- Closed quotient from the PVG (CQ_{PVG})
- Speed quotient from the kymographic contour (SQ_{kym})
- Closing quotient from the kymographic contour (CgQ_{kym})
- Vertical phase differences between the lower and upper margins of the VFs from the kymographic contour (VPD_{kym})
- Normalized amplitude quotient from the GAW (NAQ_{GA}).

In addition, the following parameters are calculated: the amplitude of the EGG signal, the sound pressure level (SPL), the spectral balance (SB), and the difference in the levels of the first and second harmonics L_{H1}-L_{H2}– see paper **III** for details.

2.4.3.Results

On average, the outsiders were able to correctly identify the registers, as intended by the singer, in 64% of the cases, and the insiders in 89% of the cases. Results from the visual analysis of the high-speed videos are summarized in Figure 33. They revealed laryngeal and physiological differences separating the low-, middle-, and high-pitched phonation regions (indicated by R_{Lo}, R_{Mid}, and R_{Hi}, respectively). Importantly, the values of NAQ_{GA} and of the CgQ_{kym} were consistently lower in chest-like register throughout the entire range (except at B5 for the CgQ_{kym}), as shown in Figure 34. This indicates that the closing phase was consistently shorter in chest-like register.

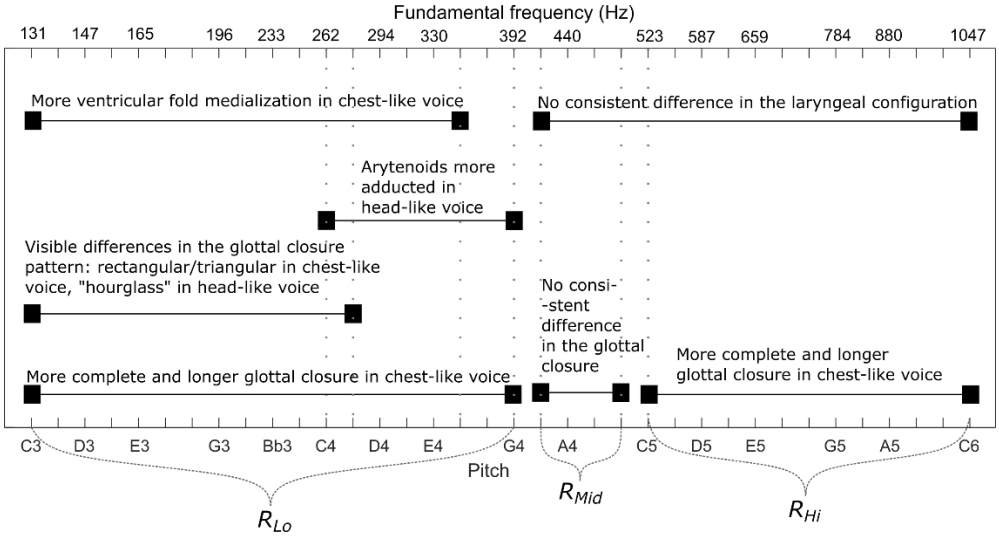


Figure 33. Diagram summarizing the visual and objective differences observed between the chest-like and head-like phonations in the high-speed videos and the corresponding ranges where those differences appeared. R_{Lo}, R_{Mid} and R_{Hi} indicate the low, middle, and high range, respectively.

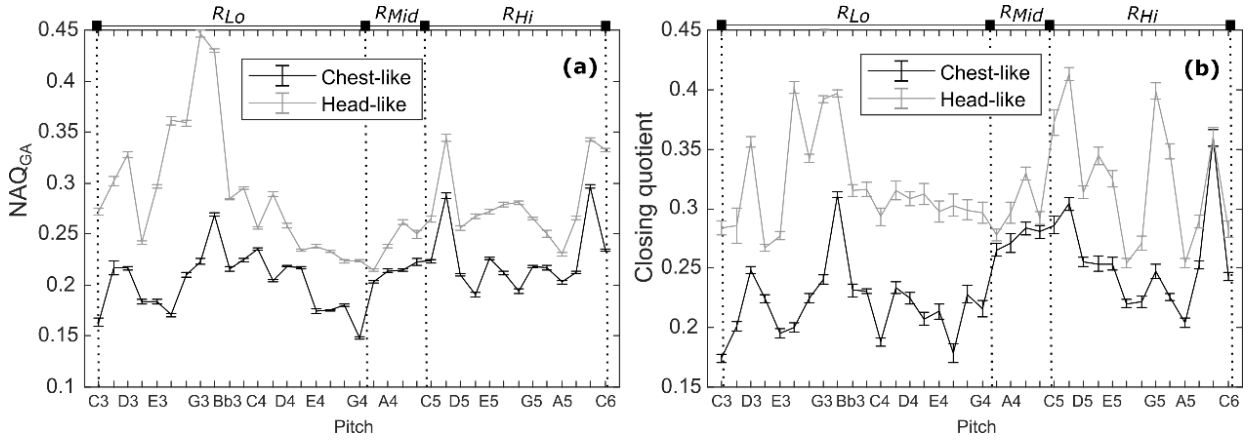


Figure 34. Mean Normalized Amplitude Quotient NAQ_{GA} (a) and Closing Quotient CgQ_{kym} (b) distinguishing the chest-like and head-like registers for every pitch. The error bars represent the standard error of the mean.

2.4.4.Discussion and conclusion

NAQ_{GA} and the CgQ_{kym} were found to be the only two parameters with consistently lower values in chest-like register across the entire pitch range (except at B5 for the CgQ_{kym}), which suggests the glottal closing speed is one of the most relevant parameters to discriminate between chest-like and head-like registers in our singer. This supports the hypothesis mentioned in section 1.4.2 that the increased activity of the TA muscle thickens the VFs and therefore increases the

vibrational amplitude of the lower margin of the VFs, which in turn increases the closing speed. Interestingly, the CQ_{PVG} was not able to distinguish the two registers across the large pitch range (not shown here for brevity – see paper **III** for details). This conflicts with the results from previous studies that mostly used CQ and OQ for differentiating registers (Sundberg et al., 2001; Roubeau et al., 2009; Echternach, Burk, Köberlein, Herbst, et al., 2017; Echternach, Burk, Köberlein, Selamtzis, et al., 2017).

Perceptually, the lower score of the outsiders group in the listening tests indicates that those listeners had more difficulties than the insiders group to recognize the registers. This suggests that the distinction between the two registers intended by the singer is not easy to recognize. As expected, the insiders group performed better, as they were familiar with the study design.

*2.5.Paper **IV**: Simulation of laryngoscopic high-speed videos from a kinematic model*

2.5.1.Objectives

Clinicians commonly use laryngoscopic imaging techniques to assess the general VF health and the presence of pathologies. Particularly, strobolaryngoscopy, and to some extent, in some places, also videokymography (VKG), are routinely used, as they provide real-time feedback and their clinical value has been demonstrated (Casiano et al., 1992; Phadke et al., 2017). Laryngeal HSV, however, requires the recorded images to be reviewed and analyzed and is therefore slow to be routinely used in clinical practice (see also the reasons described in section 1.2.1.3), even though it is the technique that can provide the most detailed information about the VFs. As such, the relation between visual features observed in laryngeal HSV images and kinematic properties of the VFs is under researched, and there is a need to improve our interpretation of laryngeal HSV. This study proposes to address this need by extending a kinematic mucosal wave model of the VF vibrations to enable the generation of synthetic HSV of the entire glottis.

2.5.2.Methods

The mucosal wave model used in this study is described in section 1.6. This model was originally designed to simulate two-dimensional (2D) vibrations, by considering a coronal (vertical left-right) slice of the VFs. Here, it is extended to three-dimensional (3D) vibrations by concatenating 256 slices along the anterior-posterior axis. Kymograms generated by each slice are then combined, to form a simulated high-speed video, where the entire glottis is visible. Geometric and kinematic parameters of the VFs are then varied along the anterior-posterior axis, to display anterior-posterior differences typically observed in vivo through laryngeal HSV. Parameters were adjusted empirically so that the simulated videos resemble reference laryngeal high-speed videos, obtained from the Vienna database of pathological and non-pathological voices (Aichinger et al., 2016).

2.5.3.Results

For brevity, this section only describes two examples of HSV simulated by the model – see paper **IV** for details. The first example is displayed in Figure 35 and corresponds to a regular pressed phonation. For this example, the pressed quality is simulated by setting a negative glottal half-width across the entire glottis, except for the most posterior position where the glottal half-width is set to 0. The amplitude of vibration is also set to 0 at the posterior position, so that the left and right VFs are always only just touching each other at this position (this creates the impression of a slight posterior glottal gap in Figure 35c due to the roundedness of the VF shape). During glottal closing, the mucosal wave is seen propagating across the top VF surface (Figure 35a,b).

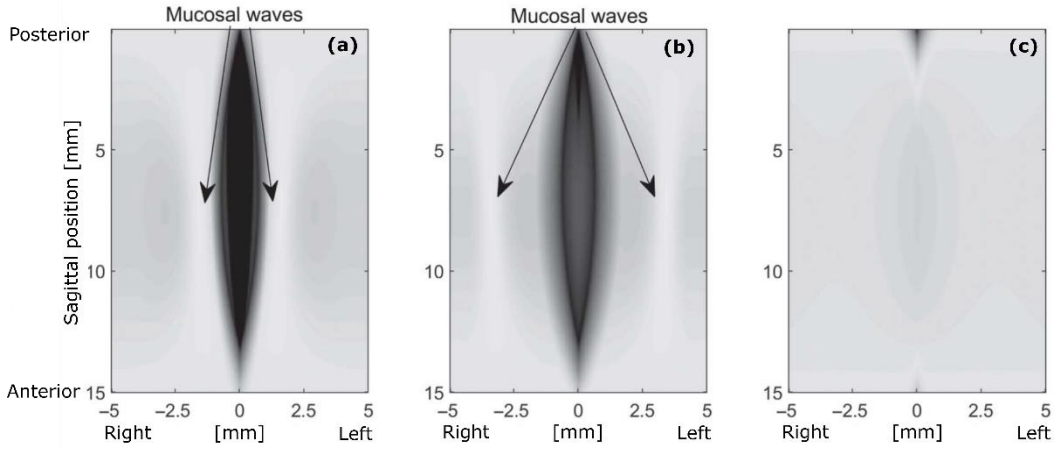


Figure 35. Illustration of pressed VF vibrations. Shown are frames of maximal opening (a), maximal closure (c), and a frame in between (b). The mucosal waves traveling laterally are also indicated by black arrows.

The second example corresponds to a pathological voice exhibiting chaotic vibrations, simulated by imposing random variations in the phase of vibration, which are different between the left and right VFs, and between their anterior and posterior part. This example is displayed through the PVG (recall section 1.2.1.4) in Figure 36. The vibration of both the left and right VFs are visibly irregular, in addition to being asymmetric.

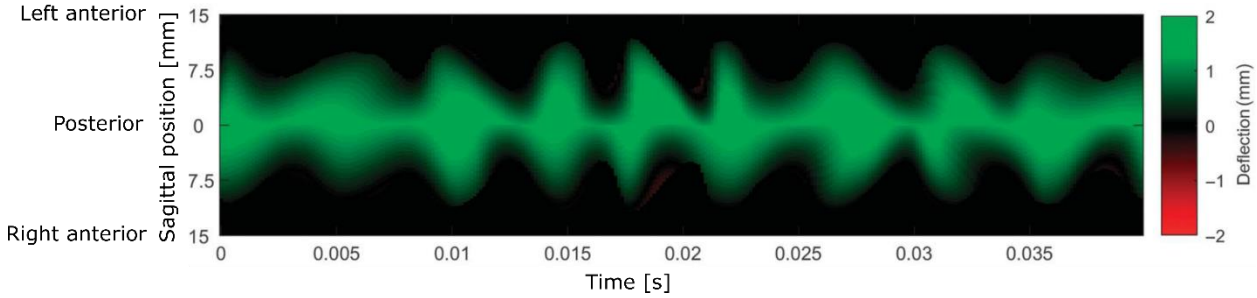


Figure 36. Phonovibrogram (PVG) corresponding to chaotic vibrations

2.5.4.Discussion and conclusion

The mucosal wave model is successfully extended to 3D, allowing the generation of synthetic videos with a realistic appearance, containing left-right and anterior-posterior differences. Additionally, the implementation of perturbations, in the form of random vibratory phase distortion, allows to synthetize irregular vibrations, which are associated with pathological voices. This includes perturbations that are different for the left and right VF, as well as for the anterior and posterior parts of the glottis. As such, this extended model offers useful insights about the effects of such perturbations in the VF kinematic parameters on the appearance of the corresponding vibrations.

3. OVERALL CONCLUSION

This thesis provides important new insights on the vibratory properties of the VFs. Specifically, the first part tests and documents a newly developed anechoic subglottal tract, to investigate the subglottal pressures and the VF vibrations in fully anechoic conditions, and to compare them with subglottally resonant conditions. It shows that the anechoic tract successfully suppresses its acoustic resonances. This subglottal tract is then used with excised deer larynges without a supraglottal tract to investigate the VF vibrations in fully anechoic conditions. In such conditions, the subglottal pressure waveform demonstrated a remarkable resemblance with the theoretical source flow waveform, which shows that the anechoic subglottal tract can be used to study the source signal without any acoustic interactions with resonance cavities and without the need for inverse filtering.

In the second part, this thesis investigates the differences in the occurrences of frequency jumps in human excised larynges, between anechoic and subglottally resonant conditions. This part presents the first experimental evidence that frequency jumps occur also in fully anechoic conditions. The results provide a proof that those frequency jumps are primarily caused by inherent nonlinear-dynamic VF properties, even though the acoustic resonances in the subglottal tract appear to play a secondary role.

The third part of this thesis investigates the relevant glottal and laryngeal parameters that can discriminate between chest and head registers throughout a wide pitch range. Here, chest and head registers are investigated over a wide pitch range in a professional female singer. Visual and objective analysis revealed that the glottal closing speed is one of the most relevant parameters to discern chest and head registers regardless of pitch, which is revealed through the consistent lower values of the normalized amplitude quotient and closing quotient in chest register. This contrasts with most previous studies that rather assumed the closed quotient to be the most sensitive measure for distinguishing the chest and head registers in singing. In our study, the closed quotient was not found to be able to distinguish the two registers as consistently as the normalized amplitude quotient and closing quotient. This part helps to advance our understanding of voice registers, particularly of the laryngeal adjustments and VF kinematics employed by trained singers to produce chest and head registers.

Finally, the fourth part aims at generating synthetic high-speed videos of the entire glottis. This was achieved through the extension of a two-dimensional existing kinematic model of the VF vibrations, referred to as the mucosal wave model here, to three dimensions. The extended model demonstrates capabilities of generating videos of the vibrating glottis that closely resemble the high-speed laryngoscopic videos obtained clinically in vivo. Particularly, the synthetic videos display left-right and anterior-posterior differences, which are typically seen in laryngeal voice disorders. Additionally, the extended model is capable of applying perturbations in the form of random phase distortions, which appear as irregular or chaotic vibrations. As such, this model could be further used to help understand the underlying mechanisms responsible for pathological voices.

4. REFERENCES

Aichinger, P., Roesner, I., Leonhard, M., Denk-Linnert, D.-M., Bigenzahn, W., & Schneider-Stickler, B. (2016). *A database of laryngeal high-speed videos with simultaneous high-quality audio recordings of pathological and non-pathological voices.* In Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16), Portorož, Slovenia.

Alberti, P. W. (1996). The History of Laryngology: A Centennial Celebration. *Otolaryngology–Head and Neck Surgery, 114*(3), 345-354.

Alderson, R. M. (1979). *Complete handbook of voice training*. West Nyack, New York: Parker Publishing Company, Inc.

Alipour, F., Finnegan, E. M., & Jaiswal, S. (2013). Phonatory Characteristics of the Excised Human Larynx in Comparison to Other Species. *J. Voice, 27*(4), 441-447.

Alku, P. (2011). Glottal inverse filtering analysis of human voice production—A review of estimation and parameterization methods of the glottal excitation and their applications. *Sadhana - Academy Proceedings in Engineering Sciences, 36*(5), 623-650.

Alku, P., Bäckström, T., & Vilkman, E. (2002). Normalized amplitude quotient for parametrization of the glottal flow. *J. Acoust. Soc. Am., 112*(2), 701-710.

Baer, T. (1975). *Investigation of phonation using excised larynxes (Doctoral dissertation).* Massachusetts Institute of Technology, Cambridge, MA.

Baer, T., Gay, T., & Niimi, S. (1976). Control of fundamental frequency, intensity and register of phonation. *Haskins Laboratories Status Report on Speech Research, 45*(46), 175-185.

Baer, T., Gore, J. C., Gracco, L. C., & Nye, P. W. (1991). Analysis of vocal tract shape and dimensions using magnetic resonance imaging: Vowels. *J. Acoust. Soc. Am., 90*(2), 799-828.

Baken, R. J. (1992). Electroglottography. *J. Voice, 6*(2), 98-110.

Baken, R. J., & Orlikoff, R. F. (2000). *Clinical measurement of speech and voice*: Cengage Learning.

Berke, G., Mendelsohn, A. H., Scott Howard, N., & Zhang, Z. (2013). Neuromuscular induced phonation in a human ex vivo perfused larynx preparation. *J. Acoust. Soc. Am., 133*(2), EL114-EL117.

Berry, D. A. (2001). Mechanisms of modal and nonmodal phonation. *J. Phon., 29*(4), 431-450.

Berry, D. A., Herzel, H., Titze, I. R., & Krischer, K. (1994). Interpretation of biomechanical simulations of normal and chaotic vocal fold oscillations with empirical eigenfunctions. *J. Acoust. Soc. Am., 95*(6), 3595-3604.

Berry, D. A., Herzel, H., Titze, I. R., & Story, B. H. (1996). Bifurcations in excised larynx experiments. *J. Voice, 10*(2), 129-138.

Berry, D. A., Montequin, D. W., & Tayama, N. (2001). High-speed digital imaging of the medial surface of the vocal folds. *J. Acoust. Soc. Am., 110*(5), 2539-2547.

Berry, D. A., & Titze, I. R. (1996). Normal modes in a continuum model of vocal fold tissues. *J. Acoust. Soc. Am., 100*(5), 3345-3354.

Björklund, S., & Sundberg, J. (2016). Relationship between subglottal pressure and sound pressure level in untrained voices. *J. Voice, 30*(1), 15-20.

Botros, A. (2001). Frequency to Musical Note Converter. Retrieved from https://newt.phys.unsw.edu.au/music/note/

Bouhuys, A., Mead, J., Proctor, D. F., & Stevens, K. N. (1968). Pressure-flow events during singing. *Ann. N. Y. Acad. Sci., 155*(1), 165-176.

Camacho, A., & Harris, J. G. (2008). A sawtooth waveform inspired pitch estimator for speech and music. *J. Acoust. Soc. Am., 124*(3), 1638-1652.

Casiano, R. R., Zaveri, V., & Lundy, D. S. (1992). Efficacy of Videostroboscopy in the Diagnosis of Voice Disorders. *Otolaryngology–Head and Neck Surgery, 107*(1), 95-100.

Castellengo, M., Lamesch, S., & Henrich, N. (2007). *Vocal registers and laryngeal mechanisms, a case study: The french «voix mixte».* In Proceedings of the 19th ICA, Madrid.

Chan, R. W., & Titze, I. R. (2003). Effect of Postmortem Changes and Freezing on the Viscoelastic Properties of Vocal Fold Tissues. *Annals of Biomedical Engineering, 31*(4), 482-491.

Collyer, S., Davis, P. J., Thorpe, C. W., & Callaghan, J. (2007). Sound pressure level and spectral balance linearity and symmetry in the messa di voce of female classical singers. *J. Acoust. Soc. Am., 121*(3), 1728-1736.
Collyer, S., Davis, P. J., Thorpe, C. W., & Callaghan, J. (2009). Fundamental frequency influences the relationship between sound pressure level and spectral balance in female classically trained singers. *J. Acoust. Soc. Am., 126*(1), 396-406.
Colton, R. H. (1972). Spectral characteristics of the modal and falsetto registers. *Folia Phoniatr. Logop., 24*(5-6), 337-344.
Cranen, B., & Boves, L. (1987). On subglottal formant analysis. *J. Acoust. Soc. Am., 81*(3), 734-746.
Deliyski, D. D., Hillman, R. E., & Mehta, D. D. (2015). Laryngeal High-Speed Videoendoscopy: Rationale and Recommendation for Accurate and Consistent Terminology. *J. Speech Lang. Hear. Res., 58*(5), 1488-1492.
Deliyski, D. D., Powell, M. E. G., Zacharias, S. R. C., Gerlach, T. T., & de Alarcon, A. (2015). Experimental Investigation on Minimum Frame Rate Requirements of High-Speed Videoendoscopy for Clinical Voice Assessment. *Biomed Signal Process Control, 17*, 21-28.
Döllinger, M., & Berry, D. A. (2006). Visualization and quantification of the medial surface dynamics of an excised human vocal fold during phonation. *J. Voice, 20*(3), 401-413.
Döllinger, M., Kobler, J., Berry, D. A., Mehta, D. D., Luegmair, G., & Bohr, C. (2011). Experiments on Analysing Voice Production: Excised (Human, Animal) and In Vivo (Animal) Approaches. *Curr. Bioinform., 6*(3), 286-304.
Döllinger, M., Lohscheller, J. r., McWhorter, A., & Kunduk, M. (2009). Variability of normal vocal fold dynamics for different vocal loading in one healthy subject investigated by phonovibrograms. *J. Voice, 23*(2), 175-181.
Durham, P. L., Scherer, R. C., Druker, D. G., & Titze, I. R. (1987). Development of excised larynx procedures for studying mechanisms of phonation. Technical report: Voice Acoustics and Biomechanics Laboratory, Department of Speech Pathology and Audiology, The University of Iowa, IA, USA.
Echternach, M., Burk, F., Köberlein, M., Herbst, C. T., Döllinger, M., Burdumy, M., & Richter, B. (2017). Oscillatory characteristics of the vocal folds across the tenor passaggio. *J. Voice, 31*(3), 381.e385-381.e314.
Echternach, M., Burk, F., Köberlein, M., Selamtzis, A., Döllinger, M., Burdumy, M., . . . Herbst, C. T. (2017). Laryngeal evidence for the first and second passaggio in professionally trained sopranos. *PLoS One, 12*(5), e0175865.
Eysholdt, U., Tigges, M., Wittenberg, T., & Pröschel, U. (1996). Direct Evaluation of High-Speed Recordings of Vocal Fold Vibrations. *Folia Phoniatr. Logop., 48*(4), 163-170.
Fabre, P. (1957). Un procédé électrique percutané d'inscrition de l'accolement glottique au cours de la phonation: glottographie de haute fréquence. Premiers résultats. *Bull. Acad. Natl. Med., 141*, 66.
Fant, G. (1960). *Acoustic Theory of Speech Production* Mouton, The Hague.
Fant, G., Ishizaka, K., Lindqvist, J., & Sundberg, J. (1972). Subglottal formants. *STL-QPSR, 1*(1972), 1-12.
Farnsworth, D. W. (1940). High-speed motion pictures of the human vocal cords. *Bell Laboratories Record, 18*, 203-208.
Ferrein, A. (1741). De la formation de la voix de l'homme. *Histoire de l'Académie Royale des Sciences, 1*, 409-432.
Fitch, W. T., & Giedd, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *J. Acoust. Soc. Am., 106*(3), 1511-1522.
Flanagan, J. L. (1972). *Speech analysis: Synthesis and perception, 2nd ed*. Oxford, England: Springer-Verlag.
Flanagan, J. L., & Landgraf, L. L. (1968). Self-Oscillating Source for Vocal-Tract Synthesizers. *IEEE Transactions on Audio and Electroacoustics, AU16*(1), 57-&.
Garcia, M. (1847). *Mémoire sur la voix humaine présenté à l'Académie des Sciences en 1840*: Duverger.
Garcia, M., & Herbst, C. T. (2018). Excised larynx experimentation: history, current developments, and prospects for bioacoustic research. *Anthropol. Sci., 126*(1), 9-17.
Garnier, M., Henrich, N., Crevier-Buchman, L., Vincent, C., Smith, J., & Wolfe, J. (2012). Glottal behavior in the high soprano range and the transition to the whistle register. *J. Acoust. Soc. Am., 131*(1), 951-962.
Grillo, H. C., Dignan, E. F., Miura, T., & Scannell, J. G. (1964). Extensive resection and reconstruction of mediastinal trachea without prosthesis or graft: an anatomical study in man. *The Journal of Thoracic and Cardiovascular Surgery, 48*(5), 741-749.
Hampala, V., Garcia, M., Švec, J. G., Scherer, R. C., & Herbst, C. T. (2016). Relationship between the electroglottographic signal and vocal fold contact area. *J. Voice, 30*(2), 161-171.

Hampala, V., Švec, J. G., Schovánek, D., & Mandát, D. (2013). Užitný vzor č. 25585: Model subglotického traktu. [Utility model no. 25585: Model of subglottal tract.] Soukup, P. 2013-27834(CZ 25505 U1), 1-7. 24-6-2013. Praha, Úřad průmyslového vlastnictví. [Prague, Czech Republic, Industrial property office].

Henrich Bernardoni, N. (2006). Mirroring the voice from Garcia to the present day: some insights into singing voice registers. *Logoped. Phoniatr. Vocol., 31*(1), 3-14.

Henrich Bernardoni, N., D'Alessandro, C., Doval, B., & Castellengo, M. (2004). On the use of the derivative of electroglottographic signals for characterization of nonpathological phonation. *J. Acoust. Soc. Am., 115*(3), 1321-1332.

Henrich Bernardoni, N., d'Alessandro, C., Doval, B., & Castellengo, M. (2005). Glottal open quotient in singing: Measurements and correlation with laryngeal mechanisms, vocal intensity, and fundamental frequency. *J. Acoust. Soc. Am., 117*(3), 1417-1430.

Henrich Bernardoni, N., Roubeau, B., & Castellengo, M. (2003). *On the use of electroglottography for characterisation of the laryngeal mechanisms.* In Proceedings of SMAC 2003, Stockholm Music Acoustics Conference, Stockholm, Sweden.

Herbst, C. T. (2019). Electroglottography – An Update. *J. Voice*.

Herbst, C. T., Elemans, C. P. H., Tokuda, I. T., Chatziioannou, V., & Švec, J. G. (2023). Dynamic System Coupling in Voice Production. *J. Voice*.

Herbst, C. T., Hampala, V., Garcia, M., Hofer, R., & Švec, J. G. (2017). Hemi-laryngeal Setup for Studying Vocal Fold Vibration in Three Dimensions. *J. Vis. Exp.*(129).

Herbst, C. T., Lohscheller, J., Švec, J. G., Henrich, N., Weissengruber, G., & Fitch, W. T. (2014). Glottal opening and closing events investigated by electroglottography and super-high-speed video recordings. *J. Exp. Biol., 217*(6), 955-963.

Herbst, C. T., Nishimura, T., Garcia, M., Migimatsu, K., & Tokuda, I. T. (2020). Effect of Ventricular Folds on Vocalization Fundamental Frequency in Domestic Pigs (Sus scrofa domesticus). *J. Voice*.

Herbst, C. T., Qiu, Q., Schutte, H. K., & Švec, J. G. (2011). Membranous and cartilaginous vocal fold adduction in singing. *J. Acoust. Soc. Am., 129*(4), 2253-2262.

Herbst, C. T., & Švec, J. G. (2014). Adjustment of glottal configurations in singing. *J. Sing., 70*, 301-308.

Hertegård, S. (2005). What have we learned about laryngeal physiology from high-speed digital videoendoscopy? *Curr Opin Otolaryngol Head Neck Surg, 13*(3), 152-156.

Herzel, H., Berry, D. A., Titze, I. R., & Saleh, M. (1994). Analysis of Vocal Disorders With Methods From Nonlinear Dynamics. *J. Speech Lang. Hear. Res., 37*(5), 1008-1019.

Hirano, M. (1974). Morphological structure of the vocal cord as a vibrator and its variations. *Folia Phoniatr. Logop., 26*(2), 89-94.

Hirano, M. (1975). Phonosurgery: basic and clinical investigations. *Otologia (Fukuoka), 21 (suppl. 1)*, 239-440.

Hirano, M. (1981). *Clinical examination of voice* (Vol. 5): Spring-Verlag.

Hirano, M., Vennard, W., & Ohala, J. (1970). Regulation of register, pitch and intensity of voice. An electromyographic investigation of intrinsic laryngeal muscles. *Folia Phoniatr. Logop., 22*(1).

Hiroto, I. (1966). The mechanism of phonation: its pathophysiological aspects. (in Japanese). *Nippon Jibiinkoka Gakkai Kaiho, 69*(12), 2097-2106.

Holbert, J. M., & Strollo, D. C. (1995). Imaging of the Normal Trachea. *Journal of Thoracic Imaging, 10*(3).

Horáček, J., & Švec, J. G. (2002). Aeroelastic model of vocal-fold-shaped vibrating element for studying the phonation threshold. *J. Fluids Struct., 16*(7), 931-955.

Horáček, J., Švec, J. G., & Šidlof, P. (2009). *Numerical simulation of videokymographic images of self-oscillating vocal folds.* In 3rd Advanced Voice Function Assessment (AFVA) International Workshop.

Horáček, J., Švec, J. G., Veselý, J., & Vilkman, E. (2004). *Bifurcations in Excised Larynges Caused by Vocal Fold Elongation.* In Proceedings of the International Conference on Voice Physiology and Biomechanics, Marseille: Laboratory of Audio-Phonology, France.

Husson, R. (1950). *Etude des phénomènes physiologiques et acoustiques fondamentaux de la voix chantée [Study of the fundamental physiologic and acoustic phenonema of the singing voice] (Doctoral thesis).* Faculté des Sciences, Éditions de la revue scientifique, Paris.

Ishizaka, K., & Flanagan, J. L. (1972). Synthesis of Voiced Sounds From a Two-Mass Model of the Vocal Cords. *Bell Syst. tech., 51*(6), 1233-1268.
Jiang, J. J., Chang, C. I., Raviv, J. R., Gupta, S., Banzali Jr, F. M., & Hanson, D. G. (2000). Quantitative study of mucosal wave via videokymography in canine larynges. *Laryngoscope, 110*(9), 1567-1573.
Jiang, J. J., & Titze, I. R. (1993). A methodological study of hemilaryngeal phonation. *Laryngoscope, 103*(8), 872-882.
Kallen, L. A. (1932). LARYNGOSTROBOSCOPY IN THE PRACTICE OF OTOLARYNGOLOGY. *Archives of Otolaryngology, 16*(6), 791-807.
Karakozoglou, S.-Z., Henrich, N., d'Alessandro, C., & Stylianou, Y. (2012). Automatic glottal segmentation using local-based active contours and application to glottovibrography. *Speech Commun., 54*(5), 641-654.
Kinsler, L. E., Frey, A. R., Coppens, A. B., & Sanders, J. V. (2000). *Fundamentals of acoustics (4th edition)*. New York: John Wiley & Sons, Inc.
Kist, A. M., Gómez, P., Dubrovskiy, D., Schlegel, P., Kunduk, M., Echternach, M., . . . Döllinger, M. (2021). A Deep Learning Enhanced Novel Software Tool for Laryngeal Dynamics Analysis. *J. Speech Lang. Hear. Res., 64*(6), 1889-1903.
Kist, A. M., Zilker, J., Gómez, P., Schützenberger, A., & Döllinger, M. (2020). Rethinking glottal midline detection. *Sci. Rep., 10*.
Kob, M., & Frauenrath, T. (2009). A system for parallel measurement of glottis opening and larynx position. *Biomed Signal Process Control, 4*(3), 221-228.
Kochis-Jennings, K. A., Finnegan, E. M., Hoffman, H. T., & Jaiswal, S. (2012). Laryngeal muscle activity and vocal fold adduction during chest, chestmix, headmix, and head registers in females. *J. Voice, 26*(2), 182-193.
Kochis-Jennings, K. A., Finnegan, E. M., Hoffman, H. T., Jaiswal, S., & Hull, D. (2014). Cricothyroid muscle and thyroarytenoid muscle dominance in vocal register control: preliminary results. *J. Voice, 28*(5), 652.e621-652.e629.
Krausert, C. R., Olszewski, A. E., Taylor, L. N., McMurray, J. S., Dailey, S. H., & Jiang, J. J. (2011). Mucosal Wave Measurement and Visualization Techniques. *J. Voice, 25*(4), 395-405.
Kreiman, J., Gerratt, B. R., & Antoñanzas-Barroso, N. (2007). Measures of the Glottal Source Spectrum. *J. Speech Lang. Hear. Res., 50*(3), 595-610.
Kumar, S. P., Phadke, K. V., Vydrová, J., Novozámský, A., Zita, A., Zitová, B., & Švec, J. G. (2020). Visual and Automatic Evaluation of Vocal Fold Mucosal Waves Through Sharpness of Lateral Peaks in High-Speed Videokymographic Images. *J. Voice, 34*(2), 170-178.
Kumar, S. P., & Švec, J. G. (2019). Kinematic model for simulating mucosal wave phenomena on vocal folds. *Biomed Signal Process Control, 49*, 328-337.
Lagier, A., Guenoun, D., Legou, T., Espesser, R., Giovanni, A., & Champsaur, P. (2017). Control of the glottal configuration in ex vivo human models: quantitative anatomy for clinical and experimental practices. *Surg. Radiol. Anat., 39*(3), 257-262.
Lagier, A., Legou, T., Silva, F., & Hélie, T. (2020). *Investigating phonation through excised human larynges: recent developments and ongoing works on an animated testbench*. Paper presented at the eForum Acusticum 2020.
Large, J. (1972). Towards an integrated physiologic-acoustic theory of vocal registers. *NATS Bulletin, 28*(3), 18-25.
Legou, T., Lagier, A., Silva, F., Bernardoni, N. H., Champsaur, P., & Giovanni, A. (2015). *Test bench for human excised larynx studies.* In 9th International Workshop Models and Analysis of Vocal Emissions for Biomedical Applications (MAVEBA 2015).
Lohscheller, J., & Eysholdt, U. (2008). Phonovibrogram visualization of entire vocal fold dynamics. *Laryngoscope, 118*(4), 753-758.
Lulich, S. M., Morton, J. R., Arsikere, H., Sommers, M. S., Leung, G. K. F., & Alwan, A. (2012). Subglottal resonances of adult male and female native speakers of American English. *J. Acoust. Soc. Am., 132*(4), 2592-2602.
Mehta, D. D., Deliyski, D. D., & Hillman, R. E. (2010). Commentary on Why Laryngeal Stroboscopy Really Works: Clarifying Misconceptions Surrounding Talbot's Law and the Persistence of Vision. *J. Speech Lang. Hear. Res., 53*(5), 1263-1267.
Merkel, C. L. (1863). Anatomie und Physiologie des menschlichen Stimm-und Sprachprogramms. *Antropophonik. Leipzig: Abel*.

Miller, D. G., & Schutte, H. K. (1985). *Characteristic patterns of sub-and supraglottal pressure variations within the glottal cycle.* In Transcr. XIIIth Symp. Care Prof. Voice.
Moore, P. (1991). A short history of laryngeal investigation. *J. Voice, 5*(3), 266-281.
Mörner, M., Fransson, F., & Fant, G. (1963). Voice register terminology and standard pitch. *STL-QPSR, 4*(4), 17-23.
Murtola, T., Aalto, A., Malinen, J., Aalto, D., & Vainio, M. (2018). Modal locking between vocal fold oscillations and vocal tract acoustics. *Acta Acust. united Ac., 104*(2), 323-337.
Novozámský, A., Sedlář, J., Zita, A., Šroubek, F., Flussef, J., Švec, J. G., . . . Zitová, B. (2015). *Image analysis of videokymographic data.* In 2015 IEEE International Conference on Image Processing (ICIP).
Patel, R. R., & Ternström, S. (2021). Quantitative and Qualitative Electroglottographic Wave Shape Differences in Children and Adults Using Voice Map–Based Analysis. *J. Speech Lang. Hear. Res., 64*(8), 2977-2995.
Phadke, K. V., Vydrová, J., Domagalská, R., & Švec, J. G. (2017). Evaluation of clinical value of videokymography for diagnosis and treatment of voice disorders. *Eur. Arch. Oto-Rhino-Laryngol., 274*(11), 3941-3949.
Qiu, Q., & Schutte, H. K. (2007). Real-time kymographic imaging for visualizing human vocal-fold vibratory function. *Review of Scientific Instruments, 78*(2), 024302.
Rothenberg, M. (1973). A new inverse-filtering technique for deriving the glottal air flow waveform during voicing. *J. Acoust. Soc. Am., 53*(6), 1632-1645.
Rothenberg, M. (1981). Acoustic interaction between the glottal source and the vocal tract. In *Vocal fold physiology* (Vol. 1, pp. 305-323).
Rothenberg, M. (1992). A multichannel electroglottograph. *J. Voice, 6*(1), 36-43.
Roubeau, B., Chevrie-Muller, C., & Arabia-Guidet, C. (1987). Electroglottographic study of the changes of voice registers. *Folia Phoniatr. Logop., 39*(6), 280-289.
Roubeau, B., Henrich, N., & Castellengo, M. (2009). Laryngeal Vibratory Mechanisms: The Notion of Vocal Register Revisited. *J. Voice, 23*(4), 425-438.
Sadolin, C. (2000). *Complete vocal technique*: Shout Publishing Copenhagen, Denmark.
Scherer, R. C., Druker, D. G., & Titze, I. R. (1988). Electroglottography and Direct Measurement of Vocal Fold Contact Area. In O. Fujimura (Ed.), *Vocal fold physiology: voice production, mechanisms, and functions* (pp. 279-291). New York: Raven Press Ltd.
Scherer, R. C., Shinwari, D., De Witt, K. J., Zhang, C., Kucinschi, B. R., & Afjeh, A. A. (2001). Intraglottal pressure profiles for a symmetric and oblique glottis with a divergence angle of 10 degrees. *J. Acoust. Soc. Am., 109*(4), 1616-1630.
Schlegel, P., Stingl, M., Kunduk, M., Kniesburges, S., Bohr, C., & Döllinger, M. (2019). Dependencies and Ill-designed Parameters Within High-speed Videoendoscopy and Acoustic Signal Analysis. *J. Voice, 33*(5), 811.e811-811.e812.
Schoentgen, J., & Aichinger, P. (2015). *Synthetic kymograms and glottal area waveforms in simulated non-neutral phonation.* In 9th International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications, MAVEBA 2015.
Schutte, H. K., & Miller, D. G. (1988). Resonanzspiele der Gesangsstimme in ihren Beziehungen zu supra-und subglottalen Druckverlaufen: Konsequenzen für die Stimmbildungstheorie. [Play of Resonances in the Singing Voice in the Supra- and Subglottal Pressure Changes: Consequences for the Theory of Voice Production]. *Folia Phoniatr. Logop., 40*, 65-73.
Stark, J. (1999). *Bel canto: a history of vocal pedagogy*. Toronto/Buffalo/London: University of Toronto Press.
Story, B. H., Titze, I. R., & Hoffman, E. A. (1996). Vocal tract area functions from magnetic resonance imaging. *J. Acoust. Soc. Am., 100*(1), 537-554.
Sundberg, J. (2018). Flow glottogram and subglottal pressure relationship in singers and untrained voices. *J. Voice, 32*(1), 23-31.
Sundberg, J. (2022). Objective Characterization of Phonation Type Using Amplitude of Flow Glottogram Pulse and of Voice Source Fundamental. *J. Voice, 36*(1), 4-14.
Sundberg, J., Andersson, M., & Hultqvist, C. (1999). Effects of subglottal pressure variation on professional baritone singers' voice sources. *J. Acoust. Soc. Am., 105*(3), 1965-1971.
Sundberg, J., & Högset, C. (2001). Voice source differences between falsetto and modal registers in counter tenors, tenors and baritones. *Logoped. Phoniatr. Vocol., 26*(1), 26-36.

Sundberg, J., Scherer, R., Hess, M., Müller, F., & Granqvist, S. (2013). Subglottal pressure oscillations accompanying phonation. *J. Voice, 27*(4), 411-421.
Švancara, P., Horáček, J., Martínek, T., & Švec, J. G. (2014). *Numerical simulation of videokymographic images from the results of the finite element model.* In Engineering Mechanics.
Švec, J. G., & Granqvist, S. (2018). Tutorial and Guidelines on Measurement of Sound Pressure Level in Voice and Speech. *J. Speech Lang. Hear. Res., 61*(3), 441-461.
Švec, J. G., & Schutte, H. K. (1996). Videokymography: high-speed line scanning of vocal fold vibration. *J. Voice, 10*(2), 201-205.
Švec, J. G., Schutte, H. K., Chen, J. C., & Titze, I. R. (2021). Integrative Insights into the Myoelastic-Aerodynamic Theory and Acoustics of Phonation. Scientific Tribute to Donald G. Miller. *J. Voice*.
Švec, J. G., Schutte, H. K., & Miller, D. G. (1999). On pitch jumps between chest and falsetto registers in voice: Data from living and excised human larynges. *J. Acoust. Soc. Am., 106*(3), 1523-1531.
Švec, J. G., & Šram, F. (2002). *Kymographic imaging of the vocal fold oscillations*. Paper presented at the 7th International Conference on Spoken Language Processing, Denver, Colorado.
Švec, J. G., Šram, F., & Schutte, H. K. (2007). Videokymography in Voice Disorders: What to Look For? *Ann Otol Rhinol Laryngol, 116*(3), 172-180.
Švec, J. G., Šram, F., & Schutte, H. K. (2009). Videokymography. In *The larynx, Third Edition, Volume 1* (pp. 251-271): Plural Publishing, San Diego.
Švec, J. G., Sundberg, J., & Hertegard, S. (2008). Three registers in an untrained female singer analyzed by videokymography, strobolaryngoscopy and sound spectrography. *J. Acoust. Soc. Am., 123*(1), 347-353.
Ternström, S. (2019). Normalized time-domain parameters for electroglottographic waveforms. *J. Acoust. Soc. Am., 146*(1), EL65-EL70.
Titze, I. R. (1980). Comments on the myoelastic-aerodynamic theory of phonation. *J. Speech Lang. Hear. Res., 23*(3), 495-510.
Titze, I. R. (1988a). A framework for the study of vocal registers. *J. Voice, 2*(3), 183-194.
Titze, I. R. (1988b). The physics of small-amplitude oscillation of the vocal folds. *J. Acoust. Soc. Am., 83*(4), 1536-1552.
Titze, I. R. (1989). A four-parameter model of the glottis and vocal fold contact area. *Speech Commun., 8*(3), 191-201.
Titze, I. R. (1990). Interpretation of the electroglottographic signal. *J. Voice, 4*(1), 1-9.
Titze, I. R. (2000a). Basic anatomy of the larynx. In *Principles of Voice Production (second printing)* (pp. 1-22): National Center for Voice and Speech Iowa City IA 52242.
Titze, I. R. (2000b). *Principles of Voice Production (second printing)*: National Center for Voice and Speech Iowa City IA 52242.
Titze, I. R. (2000c). The Source-filter theory of vowels. In *Principles of Voice Production (second printing)* (pp. 149-184): National Center for Voice and Speech Iowa City IA 52242.
Titze, I. R. (2000d). Vocal fold oscillation. In *Principles of Voice Production (second printing)* (pp. 87-122): National Center for Voice and Speech Iowa City IA 52242.
Titze, I. R. (2006a). Experiments with excised larynges. In *The myoelastic aerodynamic theory of phonation* (pp. 1-62). Denver CO and Iowa City IA: National Center for Voice and Speech.
Titze, I. R. (2006b). *The Myoelastic Aerodynamic Theory of Phonation*. Iowa City lA 52242: National Center for Voice and Speech.
Titze, I. R. (2006c). Theoretical analysis of maximum flow declination rate versus maximum area declination rate in phonation. *J. Speech Lang. Hear. Res., 49*(2), 439-447.
Titze, I. R. (2008). Nonlinear source-filter coupling in phonation: theory. *J. Acoust. Soc. Am., 123*(5), 2733-2749.
Titze, I. R., Baken, R. J., Bozeman, K. W., Granqvist, S., Henrich, N., Herbst, C. T., . . . Wolfe, J. (2015). Toward a consensus on symbolic notation of harmonics, resonances, and formants in vocalization. *J. Acoust. Soc. Am., 137*(5), 3005-3007.
Titze, I. R., Jiang, J., & Drucker, D. G. (1988). Preliminaries to the body-cover theory of pitch control. *J. Voice, 1*(4), 314-319.
Titze, I. R., Luschei, E. S., & Hirano, M. (1989). Role of the thyroarytenoid muscle in regulation of fundamental frequency. *J. Voice, 3*(3), 213-224.

Titze, I. R., Riede, T., & Popolo, P. (2008). Nonlinear source–filter coupling in phonation: Vocal exercises. *J. Acoust. Soc. Am., 123*(4), 1902-1915.
Tokuda, I. T., Horáček, J., Švec, J. G., & Herzel, H. (2007). Comparison of biomechanical modeling of register transitions and voice instabilities with excised larynx experiments. *J. Acoust. Soc. Am., 122*(1), 519-531.
Tokuda, I. T., Horáček, J., Švec, J. G., & Herzel, H. (2008). Bifurcations and chaos in register transitions of excised larynx experiments. *Chaos Interdiscip. J. Nonlinear Sci., 18*(1), 013102.
Tokuda, I. T., Zemke, M., Kob, M., & Herzel, H. (2010). Biomechanical modeling of register transitions and the role of vocal tract resonators. *J. Acoust. Soc. Am., 127*(3), 1528-1536.
Van den Berg, J. (1958). Myoelastic-aerodynamic theory of voice production. *J. Speech Lang. Hear. Res., 1*, 227-244.
Van den Berg, J., & Tan, T. S. (1959). Results of Experiments with Human Larynxes. *Pract. Oto-Rhino-Laryng., 21*(6), 425-450.
Van den Berg, J., Vennard, W., Burger, D., & Shervanian, C. C. (Writers). (1960). Voice production. The vibrating larynx. (Instructional film). In. Groningen, the Netherlands: University of Groningen.
Veltrup, R., Angerer, S., Semmler, M., Kist, A. M., Zillig, T., Zilker, J., . . . Döllinger, M. (2021). *Multi-Modal Ex Vivo Setup For 3d Imaging Of The Medial And Superior Vocal Fold Surfaces In Hemi Larynges.* In 14th International Conference Advances In Quantitative Laryngology, Voice And Speech Research (AQL).
Verdolini, K., & Titze, I. R. (1994). The Application of Laboratory Formulas to Clinical Voice Management. *National Center for Voice and Speech Status and Progress Report, 7*, 197-205.
Verikas, A., Uloza, V., Bacauskiene, M., Gelzinis, A., & Kelertas, E. (2009). Advances in laryngeal imaging. *Eur. Arch. Oto-Rhino-Laryngol., 266*(10), 1509-1520.
Vilkman, E. (1987). An Apparatus for Studying the Role of the Cricothyroid Articulation in the Voice Production of Excised Human Larynges. *Folia Phoniatr. Logop., 39*(4), 169-177.
Vorperian, H. K., Wang, S., Chung, M. K., Schimek, E. M., Durtschi, R. B., Kent, R. D., . . . Gentry, L. R. (2009). Anatomic development of the oral and pharyngeal portions of the vocal tract: An imaging study. *J. Acoust. Soc. Am., 125*(3), 1666-1678.
Wade, L., Hanna, N., Smith, J., & Wolfe, J. (2016). *Soprano singing, with and without resonances.* In Proceedings of Meetings on Acoustics.
Wade, L., Hanna, N., Smith, J., & Wolfe, J. (2017). The role of vocal tract and subglottal resonances in producing vocal instabilities. *J. Acoust. Soc. Am., 141*(3), 1546.
Webb, M. (2008). Mal Webb Sideways Yodeling. Retrieved from https://www.youtube.com/watch?v=8_6nNWX7TTI
Wolfe, J. (2001). Note names, MIDI numbers and frequencies. Retrieved from https://newt.phys.unsw.edu.au/jw/notes.html
Wolfe, J., Tze Wei Chu, D., Chen, J.-M., & Smith, J. (2016). An Experimentally Measured Source–Filter Model: Glottal Flow, Vocal Tract Gain and Output Sound from a Physical Model. *Acoust. Aust., 44*(1), 187-191.
Wu, L., & Zhang, Z. (2019). Voice production in a MRI-based subject-specific vocal fold model with parametrically controlled medial surface shape. *J. Acoust. Soc. Am., 146*(6), 4190-4198.
Yamauchi, A., Yokonishi, H., Imagawa, H., Sakakibara, K.-I., Nito, T., Tayama, N., & Yamasoba, T. (2015). Quantitative Analysis of Digital Videokymography: A Preliminary Study on Age- and Gender-Related Difference of Vocal Fold Vibration in Normal Speakers. *J. Voice, 29*(1), 109-119.
Zañartu, M., Mehta, D. D., Ho, J. C., Wodicka, G. R., & Hillman, R. E. (2011). Observation and analysis of in vivo vocal fold tissue instabilities produced by nonlinear source-filter coupling: a case study. *J. Acoust. Soc. Am., 129*(1), 326-339.
Zemlin, W. R. (1981). Phonation. In W. R. Zemlin (Ed.), *Speech and hearing science: anatomy and physiology (second edition)* (pp. 100-196). Englewood Cliffs: Prentice-Hall, Inc.
Zhang, Z. (2016). Cause-effect relationship between vocal fold physiology and voice production in a three-dimensional phonation model. *J. Acoust. Soc. Am., 139*(4), 1493-1507.
Zhang, Z., Neubauer, J., & Berry, D. A. (2006). The influence of subglottal acoustics on laboratory models of phonation. *J. Acoust. Soc. Am., 120*(3), 1558-1569.
Zita, A., Novozámský, A., Zitová, B., Šorel, M., Herbst, C. T., Vydrová, J., & Švec, J. G. (2022). Videokymogram Analyzer Tool: Human–computer comparison. *Biomed Signal Process Control, 78*, 103878.