Investigations of the acoustics of the vocal tract and vocal folds in vivo, ex vivo and in vitro

Noël Hanna

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

préparée dans le cadre d'une cotutelle entre l'*Université Grenoble Alpes* et The University of New South Wales, Australia

Spécialité : **Ingénierie de la cognition, de l'interaction, de l'apprentissage et de la création**

Arrêté ministériel : le 6 janvier 2005 - 7 août 2006

Présentée par

Noël Nagy Nabih HANNA

Thèse dirigée par **Nathalie HENRICH BERNARDONI** et **Joe WOLFE**
codirigée par **John SMITH**

préparée au sein du **Departement Parole et Cognition (DPC) de GIPSA-lab et le laboratoire Acoustics**

dans **l'École Doctorale Ingénierie pour La Santé, La Cognition et l'Environnement (EDISCE) et School of Physics, The University of New South Wales**

Etude aero-acoustique du conduit vocal et des plis vocaux *in vivo*, *ex vivo* et *in vitro*

Thèse soutenue publiquement le **17 Décembre 2014**,
devant le jury composé de :

Professor, Sten, TERNSTRÖM
Department of Speech, Music and Hearing, Kungliga Tekniska Högskolan (Président et Rapporteur)
Professor, Neville, FLETCHER
College of Physical and Mathematical Sciences, Australian National University (Rapporteur)
Professor, Malte, KOB
Detmold Hochschule für Musik (Membre)
Chargé de Recherche, Nathalie, HENRICH BERNADONI
CNRS, Gipsa-lab (Membre)
Professor, Joe, WOLFE
School of Physics, The University of New South Wales (Membre)
Associate Professor, John, SMITH
School of Physics, The University of New South Wales (Membre)

Investigations of the acoustics of the vocal tract and vocal folds *in vivo*, *ex vivo* and *in vitro*

Noel Nagy Nabih Hanna

A thesis in fulfilment of the requirements for the degree of

Doctor of Philosophy

School of Physics

Faculty of Science

The University of New South Wales, Australia

&

Ecole Doctorale Ingénierie pour La Santé, La Cognition et L'Envrionnement

Université de Grenoble, France

October 2014

Table of Contents

List of Symbols and Abbreviations

The following standard and non-standard symbols and abbreviations are used throughout this thesis.

$\mathcal{A}$i ($\mathcal{A}$1, $\mathcal{A}$2, …)	vocal tract anti-resonance
a, b, c	fitting parameters in chapters 5 and 6
B (B_{R1})	bandwidth (of resonance R1)
$b_{\mathcal{A}0}$, b_{R0}	resistance associated with $\mathcal{A}$0 or R0
C_t	tissue compliance
c	speed of sound
Fi (F1, F2, …)	formant
f_0	fundamental frequency of phonation
f_m	mechanical resonance frequency
k	spring constant
L_t	tissue inertance
l	length (of vocal tract)
l_g	effective glottal length
l_{sg}	effective subglottal length
m	mass of vocal tract walls
P	pressure
P_A	atmospheric pressure
P_{io}	intra-oral pressure
P_{sg}	subglottal pressure
P_w	water pressure in the latex vocal folds
Q (Q_{R1})	Quality factor (of resonance R1)
Ri (R1, R2, …)	vocal tract resonance
r	radius
r_g	effective glottal radius
r_{sg}	effective subglottal radius
S	inner surface area of the vocal tract
$S\mathcal{A}$i	impedance maximum at the lips due to subglottal tract
SRi	impedance minimum at the lips due to subglottal tract
V	volume
v	particle velocity
w	thickness of vocal tract walls
X	cross-sectional area
Z	acoustic impedance
Z_L	load acoustic impedance
Z_{vt}	acoustic impedance of the vocal tract
Z_{sg}	acoustic impedance of the subglottal tract
Z_g	acoustic impedance of the glottis
$\lvert Z\rvert$ ($\lvert Z\rvert_{R1}$)	acoustic impedance magnitude (of resonance R1)
α	attenuation coefficient
α_g	attenuation coefficient for the glottis
α_{sg}	attenuation coefficient for the subglottal tract
Γ	complex wavenumber
Γ_{sg}	wavenumber for the subglottal tract
γ	adiabatic constant
η	viscosity
λ	wavelength
ρ	density
Φ	airflow
ω	angular frequency

List of Figures

List of Tables

Acknowledgements

This thesis could not have been completed without the assistance, advice and support of many people in both Australia and France.

Above all, I thank my supervisors Joe Wolfe, John Smith and Nathalie Henrich Bernardoni, for taking me on, for educating me as an acoustician and voice researcher, and for their friendship and support.

Specific acknowledgements

Chapter 2:

Dr Craig Hardman allowed me to audit his anatomy classes and discussed with me the anatomy of the voice.

Olivia Cox produced the illustrations in Figure 2-1, Figure 2-2, Figure 2-3 and Figure 2-6.

Chapters 3 and 4:

Andrei Skougarevsky and Pritipal Baweja designed and constructed several pieces of essential apparatus, notably the experimental three microphone impedance head and the brass 'infinite' impedance calibration load.

Volunteer subjects generously gave their time to perform the experiments.

Chapter 5:

The Anatomy Laboratory of the French Alps (Laboratiore d'Anatomie des Alpes Francaises) made available the facilities for conducting the excised larynx experiments, and the anatomical pieces, from those who had donated their bodies for scientific research.

The work described in this chapter is part of an inter-disciplinary and inter-institutional collaboration, Projet Larynx, between GIPSA-lab, The Anatomy Laboratory of the French Alps (Laboratoire d'Anatomie des Alpes Françaises), the Anatomy Laboratory of CHU La Timone in Marseille, and the Speech and Language Laboratory (LPL) in Aix-en-Provence.

The setup for the excised larynx experiments was a modification of the experimental equipment used at CHU and LPL. Modifications to the experimental bench, and the dissection, mounting and control of the larynges were performed by medical students: Fahd Bennani, Adrian Mancini, Leila Morand, and Mathilde Pelloux, as part of their Masters research degrees under supervision of Philippe Chaffanjon and Yohann Robert. Laurent Ott and Ghatfan Hasan in the mechanical workshop also performed some essential modifications.

The data acquisition for larynges LF4, LM5 and LM6 was performed by Nathalie Henrich Bernardoni and Thierry Legou. Fabrice Silva assisted with the data acquisition for larynx LM1. Xavier Laval recorded the high speed video of LF4 and LM1.

Chapter 6:

The *in vitro* vocal fold replica was originally designed by Nicolas Ruty, who also performed some preliminary experiments, the results of which are discussed in this thesis.

Assistance setting up the experiment was provided by Xavier Laval.

Remerciements

It has been my pleasure to work (and play) with the Acoustics group at UNSW past and present: Jer Ming Chen, Henri Boutin, Weicong Li, Andre Almeida, Derek Chu, Kai Wen Li, David George, Jeremy Lim, Laura Wade and Andrew Skougarevsky.

I have also been warmly welcomed by the community at GIPSA-Lab, especially by my office mates: Julian Valdes Vargas, Zuheng Ming, Nicolas Hermant, Louis Delebeque, and Cedrik Erbsen.

Phillipe Masson et Nathalie Cau, je vous remercie de votre accueil, vos conseils et votre guide dans la vie pratique française, et bien sur votre patience avec ma français.

I sincerely thank the former and current heads of the School of Physics, Richard Newbury and Sven Rogge, the Postgraduate Coordinator Michael Ashley, and the Secrétariat of EDISCE Caroline Zala, for their support and patience with the cotutelle arrangement.

At home, I am lucky to have the support of my parents and parents-in-law, and most importantly, I thank my wife Natalie and daughter Ondine for their enduring love and support.

Details of funding

During the course of my PhD research I have received funding from the following sources:

Australian Postgraduate Award scholarship

The University of New South Wales, School of Physics Top Up scholarship

Region Rhone Alpes Accueil Doc Scholarship

The University of New South Wales, Postgraduate Research Student Support travel award to attend ICVPB

Two travel awards from the New South Wales division of the Australian Acoustical Society to attend the AAS Conference and the ICA

GIPSA-lab travel funding to attend PEVOC and MAVEBA

Elsevier Young Scientist Best Oral Presentation prize for the presentation of "Singing Excised Larynges: relationship between subglottal pressure and fundamental frequency" (Hanna *et al.*, 2013a) at MAVEBA

More generally, the work presented in chapters 3 and 4 was supported by Australian Research Council grant number DP1095299, and funding for the excised larynx project, of which chapter 5 is a part, was provided by CNRS through PEPS PTI 2013 (Theoretical Physics and its Interactions, Project LARYNX).

Details of Publications

Findings from research performed during the course of this degree and described in this thesis have been published as conference papers, and presented at international conferences, in collaboration with:

Professor Joe Wolfe, Associate Professor John Smith, Andrei Almeida, Jer Ming Chen and David George from the School of Physics, The University of New South Wales, Australia

Chargée de Recherche Nathalie Henrich from the Centre National de la Recherche Scientifique (CNRS), and GIPSA-lab at the Université de Grenoble, France

Ingénieur d'études Xavier Laval from GIPSA-lab at the Université de Grenoble, France

Professeur Phillipe Chaffanjon, Dr Yohann Robert, Fahd Bennani, Mathilde Pelloux, Leila Morand, and Adrian Mancini from the Laboratoire d'Anatomie Des Alpes Françaises (LADAF), Faculté de Médecine de Grenoble, France

Ingénieur de Recherche Thierry Legou from Laboratoire Parole et Langage (LPL), Aix-en-Provence, France

The articles and extended abstracts marked with an asterisk are included in Appendix B.

Journal Articles

Laryngeal control of fundamental frequency in excised human larynges, Noel Hanna, Nathalie Henrich Bernardoni, Fahd Bennani, Leila Morand, Mathilde Pelloux, Adrian Mancini, Xavier Laval, Thierry Legou, Philippe Chaffanjon (submitted 2014) (Hanna *et al.*, 2014a)

Refereed Conference Papers

*Low frequency response of the vocal tract: acoustic and mechanical resonances and their losses, Noel Hanna, John Smith and Joe Wolfe. Proceedings of Acoustics 2012 – Fremantle (Hanna *et al.*, 2012a)

*Wolfe, J., Almeida, A., Chen, J. M., George, D., Hanna, N., Smith, J., The player-wind instrument interaction, Stockholm Music Acoustics Conference (SMAC), Stockholm, Sweden, 2013 (Wolfe *et al.*, 2013)

Invited Presentations

Phonation into straws: impact of an aeroacoustical load on in-vitro vocal fold vibration, Noel Hanna, Xavier Laval, and Nathalie Henrich. Pan European Voice Conference, Prague, Czech Republic, 2013 (Hanna *et al.*, 2013b)

Other Conference Presentations and Posters

Acoustic measurements of vocal tract resonances, Noel Hanna, John Smith, and Joe Wolfe. International Conference on Voice Physiology and Biomechanics, Erlangen, Germany, 2012 (Hanna *et al.*, 2012b)

*Resonances and bandwidths in the vocal tract and why they are important for speech comprehension, Noel Hanna, John Smith, and Joe Wolfe. Australian Institute of Physics Conference, Sydney, Australia, 2012 (Hanna *et al.*, 2012c)

*Measurements of the aero-acoustic properties of the vocal folds and vocal tract during phonation into controlled acoustic loads, Noel Hanna, John Smith, and Joe Wolfe. International Congress on Acoustics, Montreal, Canada, 2013 (Hanna *et al.*, 2013c)

*Singing excised human larynges: relationship between subglottal pressure and fundamental frequency, N. Hanna, N. Henrich, A. Mancini, T. Legou, X. Laval, P. Chaffanjon, International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications (MAVEBA), Firenze, Italy, 2013 (Awarded Young Scientist Best Oral Presentation) (Hanna *et al.*, 2013a)

Singing excised human larynges: Investigating aerodynamical and biomechanical control of phonation, N. Hanna, N. Henrich, F. Bennani, L. Morand, M. Pelloux, T. Legou, P. Chaffanjon, International Conference on Voice Physiology and Biomechanics, Salt Lake City, USA, 2014 (presented by Thierry Legou) (Hanna *et al.*, 2014b)

Construction d'un modèle laryngé hybride. Mise au point d'un banc d'essai, Robert, Y., Bennani, F., Hanna, N., Henrich, N., Chaffanjon, P. Congres des Morphologistes 2014, 98, 136–137 (presented by Fahd Bennani) (Robert *et al.*, 2014)

Résumé de la thèse

La parole et le chant ont une importance capitale dans la culture humaine. Cependant, les phénomènes physiques de production et de contrôle de la voix sont encore mal compris, et leurs paramètres mal connus, principalement en raison de la difficulté d'y accéder in vivo. Dans le modèle source-filtre simplifié, la source sonore est produite par l'oscillation des plis vocaux à une fréquence fondamentale et ses multiples ; les résonances du conduit vocal filtrent l'enveloppe spectrale du signal pour produire des voyelles. Dans cette thèse, les propriétés de la source et du filtre sont étudiées et une expérience *in vitro* examine l'influence du filtre sur la source.

L'influence des paramètres de contrôle aérodynamiques ou mécaniques sur la fréquence fondamentale est étudiée *ex vivo* en utilisant des larynx humains excisés. Quatre types de discontinuités ou d'hystérésis sont observés. En dehors de ces zones de bifurcation, la fréquence fondamentale est approximativement proportionnelle à la racine carrée de la pression sous-glottique, ce qui a des implications pour le chant et de la parole, en particulier dans les langues tonales. De plus, le flux d'air qui traverse la glotte provoque un rétrécissement du conduit aryépiglottique sous l'effet de forces de pression aérodynamique, et peut initier l'oscillation des plis ventriculaires et/ou aryépiglottiques sans contrôle musculaire.

L'impédance acoustique de conduits vocaux fut mesurée *in vivo* sur un intervalle de9 octaves en fréquence et de 80 dB en amplitude, avec la glotte fermée puis pendant la phonation. Les fréquences, amplitudes et largeurs de bande des résonances acoustiques et des résonances mécaniques des tissus autour du conduit vocal sont estimées. Lorsque la glotte est fermée, les largeurs de bande et les pertes d'énergie correspondantes dans le conduit vocal sont largement supérieures aux pertes viscothermiques d'un cylindre rigide lisse, et sont encore plus importantes pendant la phonation. En utilisant un modèle simple de conduit vocal et les mesures effectuées en inspirant, des résonances acoustiques du système sous-glottique sont également estimées.

Les effets possibles de la charge aéroacoustique du filtre sur la source sont mis en évidence dans une expérience sur une maquette de plis vocaux constituée de boudins de latex remplis d'eau couplés à un tuyau rigide. La modification de la charge acoustique en aval des plis vocaux, par insertion d'une paille à l'extrémité du conduit, modifie la fréquence

fondamentale de vibration des plis. Ce résultat est discuté dans le contexte des méthodes de rééducation orthophonique à la paille couramment utilisées en thérapie de la parole.

Résumé des chapitres

Cette thèse se compose de 7 chapitres: introduction, revue de la littérature, quatre chapitres décrivant les expériences scientifiques, et une conclusion.

Les chapitres 3 et 4 présentent des mesures d'impédance acoustique du conduit vocal entre 10 Hz et 4,2 kHz obtenues en appliquant un signal acoustique à large bande entre les lèvres. Lorsque la glotte est fermée, Les résonances acoustiques du conduit avec la glotte fermée, responsables des trois premiers formants de la voix, sont situées entre 300 Hz et 4 kHz environ. et ont des amplitudes de 0,1 $MPa{\cdot}s{\cdot}m^{-3}$. Sur cet intervalle, les anti-résonances (pics d'impédance) atteignent 10 $MPa{\cdot}s{\cdot}m^{-3}$. Lorsque les lèvres sont ouvertes et en l'absence de rayonnement au niveau des lèvres, les largeurs des résonances et antirésonances acoustiques augmentent avec la fréquence, de ~50 à 90 Hz pour les hommes et de ~70 à 90 Hz pour les femmes. Ces largeurs de bande correspondent à des pertes cinq fois plus grandes que les pertes visco-thermiques d'un cylindre lisse rigide de dimensions comparables. La fonction de transfert de la pression entre les lèvres et la glotte a été estimée à l'aide d'un modèle acoustique simple. Les mesures de largeur de bandes de formants coïncident avec les estimations dans la littérature. L'impédance acoustique de la glotte ouverte est finie alors qu'elle est infiniment grande pour la glotte fermée. Au cours des mesures de phonation, l'impédance acoustique dans la glotte est inertive, ce qui augmente la largeur de bande de la résonance mais ne change pas sa fréquence.

Les mesures acoustiques et électromagnétiques révèlent la présence d'une résonance et d'une antirésonance en dessous de 300 Hz. Ces mesures permettent d'estimer l'inertance mécanique de la paroi du conduit vocal. Elle correspond à une épaisseur de paroi uniforme comprise entre 1 et 2 cm. Le premier minimum d'impédance, autour de 20 Hz, fournit un ordre de grandeur de la raideur surfacique de la paroi du conduit vocal entre 2 et 4 $kN.m^{-3}$. Le premier maximum d'impédance mesurée dans la bouche se situe autour de 200 Hz et sa fréquence augmente avec l'ouverture de la glotte. Lorsque l'ouverture des lèvres tend vers zéro, ce pic d'impédance fournit une approximation de la limite inférieure du premier formant.

Lorsque le rayon glottique efficace dépasse 1 mm, l'impédance de la glotte affecte l'impédance mesurée aux lèvres. Cette impédance est relativement bien modélisée par l'ajout d'un cylindre de longueur 20 cm ouvert à son extrémité (les poumons), qui fournit

une charge acoustique à la glotte. Lors d'une phonation chuchotée, les résonances de l'appareil sous-glottique peuvent influencer le mouvement des plis vocaux selon que leur réponse est modélisée par charge acoustique en série ou en parallèle avec la source glottique.

Le chapitre 5 décrit des expériences sur plusieurs larynx humains excisés, en l'absence de conduit vocal supraglottique. La source d'air comprimé, placée sous la glotte, fournit une charge acoustique qui peut affecter le mouvement des plis vocaux. L'oscillation de la glotte a été obtenue en modifiant la pression sous-glottique P_{sg} et la tension appliquée aux plis vocaux sur de larges intervalles. En général, l'augmentation des paramètres de contrôle aérodynamiques ou bien des paramètres mécaniques provoque un glissando vers les hautes fréquences, et leur diminution un glissando vers les basses fréquences. Quatre types de comportements discontinus ou d'hystérésis ont été rencontrés: des sauts intra- et inter-mécanisme, des changements discontinus menant à des oscillations sous-harmoniques, et des hystérésis viscoélastiques. Une relation de proportionnalité a été constatée de manière empirique entre la fréquence fondamentale f_0 et $\sqrt{P_{sg}}$.

Par ailleurs les interactions fluide-structure à l'intérieur du larynx induisent un rétrécissement du larynx aryépiglottique, et dans certains cas, les vibrations provoquent l'oscillation des plis aryépiglottiques et des plis ventriculaires. Cela met en évidence le manque de connaissance du contrôle *in vivo* du larynx aryépiglottique et de la façon dont il est modélisé *ex vivo*. Aucune preuve de l'influence de la plomberie sous-glottique sur le comportement des plis vocaux n'a été observée.

Dans le chapitre 6, un protocole expérimental utilise une maquette de glotte *in vitro* en latex. Trois pailles sont successivement introduites entre les plis vocaux en latex remplis d'eau. Cela simule un exercice d'orthophonie au cours duquel le sujet chante avec des pailles placées entre les lèvres. La technique modifie les propriétés acoustiques et aérodynamiques du conduit vocal «vu» par les plis vocaux, et affecte donc leur vibration. Dans ces expériences, la source d'air sous-glottique était choisie pour avoir peu d'influence sur l'oscillation des plis vocaux.

L'oscillation des plis vocaux de la maquette est commandée par une pression sous-glottique réglable, et ses propriétés mécaniques sont contrôlées par la pression d'eau dans les plis en latex. L'augmentation de la pression de l'eau accroît la raideur des plis et ainsi la fréquence

des résonances mécaniques. La maquette est connectée à un modèle de conduit vocal rigide, soit ouvert à l'extrémité de la glotte, soit obstrué par l'une des trois pailles.

Lorsque la gotte est ouverte, f_0 est comparable à la plus grande résonance mécanique des plis vocaux. En présence d'une paille, la pression transglottique décroit et rapproche f_0 de la résonance mécanique precedent. Ce résultat montre que l'oscillation des plis vocaux est affectée non seulement par la raideur des plis (pression hydraulique interne), mais aussi par leurs charges aérodynamique et acoustique, leurs résonances mécaniques et la pression transglottique.

Sur l'intervalle balayé par f_0, la présence d'une paille modifie également la charge acoustique appliquée au conduit vocal initialement inertive en une compliance. Cet effet secondaire n'est pas nécessairement observé avec un conduit vocal non-rigide, par exemple *in vivo*, mais il peut également influencer la vibration des plis vocaux.

Abstract

Speech and singing are of enormous importance to human culture, yet the physics that underlies the production and control of the voice is incompletely understood, and its parameters not well known, mainly due to the difficulty of accessing them *in vivo*. In the simplified but well-accepted source-filter model, non-linear vocal fold oscillation produces a sound source at a fundamental frequency and its multiples, the resonances of the vocal tract filter the spectral envelope of the sound to produce voice formants. In this thesis, both source and tract properties are studied experimentally and an *in vitro* experiment investigates how the filter can affect the source.

The control of fundamental frequency by either air supply or mechanical control parameters is investigated *ex vivo* using excised human larynges. All else equal, and excluding the four types of discontinuity or hysteresis observed, the fundamental frequency was found to be proportional to the square root of subglottal pressure, which has implications for singing and speech production, particularly in tonal languages. Additionally, airflow through the glottis causes a narrowing of the aryepiglottic tube and can initiate ventricular and/or aryepiglottic fold oscillation without muscular control.

The acoustic impedance of the vocal tract was measured *in vivo* over a range of 9 octaves and 80 dB dynamic range with the glottis closed and during phonation. The frequencies, magnitudes and bandwidths were measured for the acoustic and for the mechanical resonances of the surrounding tissues. The bandwidths and the energy losses in the vocal tract that cause them were found to be five-fold higher than the visco-thermal losses of a dry, smooth rigid cylinder, and to increase during phonation. Using a simple vocal tract model and measurements during inhalation, the subglottal system resonances were also estimated.

The possible effects of the filter on the source are demonstrated in an experiment on a water-filled latex vocal fold replica: changing the aero-acoustic load of the model tract by inserting a straw at the model lips changes the fundamental frequency. This result is discussed in the context of straw phonation used in speech therapy.

1 Introduction

Speech is of paramount importance to human communication and the voice is one of the oldest and by its nature the most widespread of musical instruments. Our common understanding of the centrality of the vocal folds in producing the sound of the voice owes a debt to Ferrein (1741), who from anatomical observation, coined the term vocal cords (*cordes vocales*) to describe the vibrating source of the voice in the larynx. Ferrein observed that these cords (now called vocal folds) were set into motion by the flow of air between them, and formulated analogies with the bowed strings of the viola de gamba, and the plucked strings of the harpsichord. Previously, established ideas of voice production were based on the written accounts c. AD 200 of the Greek physician Galen of Pergamon, who suggested both that the shape of the glottis (aperture between the vocal folds) acted like the double reed of an Aulos (a rather better analogy), but also that the exhaled air was plucked while passing through the larynx much like a stringed instrument (a summary of several translations of his work are presented in Wollock (1997)). However, the opening and closing of the vocal folds during voice production did not gain wide acceptance until direct evidence by stroboscopic laryngoscopy, presumably in the early 20th century, was presented (Russell and Cotton, 1959).

In the latter half of the 20th century understanding of voice production advanced due to Van den Berg's myoelastic-aerodynamic theory of voice production (1958) that combined the role of muscles, biomechanical tissue properties and aerodynamics to explain vocal fold vibration. Fant's source-filter theory (Fant, 1960) provided a model that is still widely used for voice analysis and synthesis. In the simplest version of this theory, the source (i.e. the acoustic current modulated by the vocal folds) and the filter (the vocal tract) are treated as independent. In this version, the source has no effect on the filter, and that the filter properties have no dependence on the source. To the first order, this assumption allows for semi-quantitative understanding of the acoustics of voice production, particularly in cases when the fundamental frequency of vibration of the vocal folds is lower than the frequency of the first acoustic resonance of the vocal tract, as is typically the case in speech.

However, even with the current level of understanding, there are important parameters that remain unknown, often due to the difficulties of making direct measurements on the vocal folds and vocal tract while in use. The following comment from one of the pioneers on the state of research in 2001 is still apt today.

Speech research is facing a challenge. Without a broad integrated expansion of fundamental knowledge we cannot realise far reaching aims such as communicating with computers as freely as with human beings. (Fant, 2001)

The aim of this thesis is to contribute to this expansion of fundamental knowledge through a detailed *in vivo*[1] study of the vocal tract, *ex vivo* experiments on the vocal folds, and *in vitro* tests on the limitations of the source-filter model of speech production. These experiments address the following open questions in voice research:

> What are the magnitudes and bandwidths of the supraglottal vocal tract resonances, both acoustic and mechanical?
>
> What are the frequencies of the subglottal resonances?
>
> What is the relationship between subglottal pressure and fundamental frequency?
>
> How does laryngeal control affect the fundamental frequency?
>
> Does straw phonation show evidence of source-filter interaction?

The thesis is divided into seven chapters, including this very general introduction. Chapter 2 provides a brief overview of the anatomy and physics of the voice and the current state of knowledge of voice production, with particular reference to the source-filter model. Chapter 3 introduces the three-microphone-three-calibration acoustic impedance measurement technique adapted to measure the resonances of the *in vivo* vocal tract with a closed glottis. With the aid of a simple acoustic model, the measured frequencies, amplitudes and bandwidths are related to the energy losses and mechanical properties of a simple acoustic duct. In Chapter 4 this *in vivo* measurement technique is extended to analyse the influence of the glottis and subglottal vocal tract on the impedance measured at the mouth, and the acoustic model is extended to include the subglottal tract. In Chapter 5 the relationships between aerodynamic and mechanical laryngeal control parameters and the fundamental frequency of the vocal folds are investigated in *ex vivo* experiments on excised human larynges. The experiments show that the fundamental frequency of phonation is empirically approximated as being proportional to the square root of subglottal pressure with both tested methods of laryngeal control. Chapter 6 describes a series of *in vitro* experiments that simulate the straw phonation technique used in speech

[1] The Latin terms *in vivo*, *ex vivo*, *in vitro* are used throughout this thesis to refer to experiments performed; on living subjects, using human anatomical specimens excised from cadavers, and with synthetic replicas, respectively.

therapy. These demonstrate a type of source-filter interaction in which the downstream acoustic load provided by the straw affects the vibration of the water-filled latex vocal fold replica. Because of the diverse experimental methods described in this thesis, each of the experimental chapters contains its own Materials and Methods section. A summary of the experimental results and concluding remarks are presented in Chapter 7.

2 Background & Literature Review

2.1. Anatomy and physiology of the vocal apparatus

The human vocal apparatus is not an independent system in the anatomical sense; instead it shares elements with both the respiratory and digestive systems, and consists of a range of organs, muscles, cartilages and airways, extending from the diaphragm via the larynx (or more colloquially the voice box), to the air at the lips and nose, as shown in Figure 2-1. The production of voiced sounds, henceforth referred to as phonation[2], requires precise coordination of respiration, control of laryngeal muscles, and positioning of the tongue, lips and jaw.

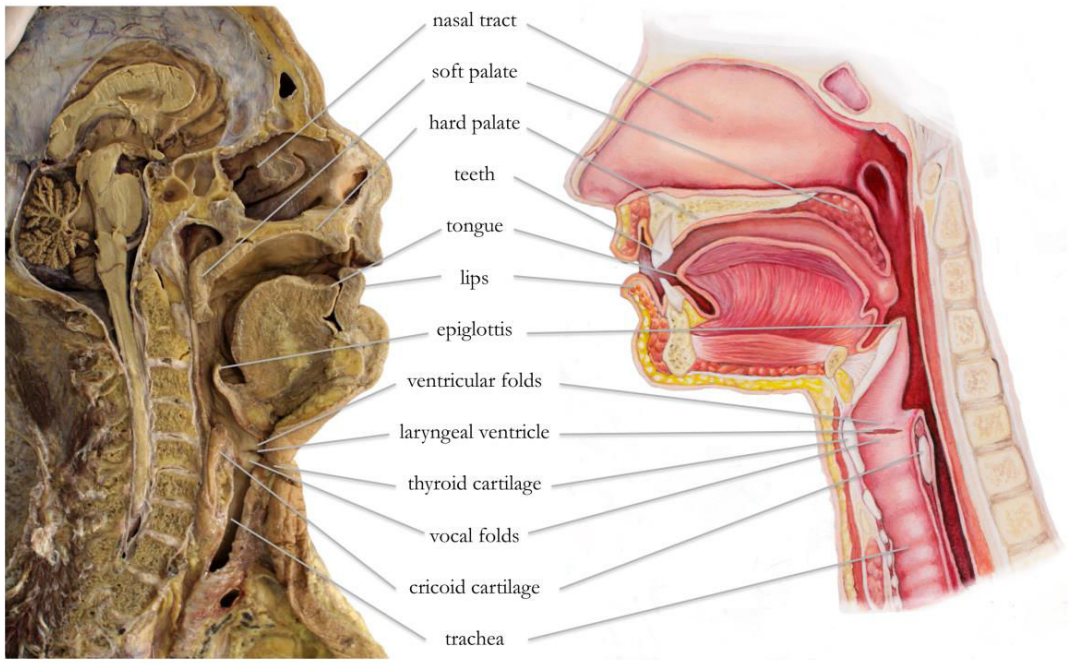


Figure 2-1. Side views of the vocal apparatus from a photograph of a preserved cadaver (left) and a simplified illustration (right). Parts of the anatomy of relevance to this thesis are labelled. Illustration by Olivia Cox.

From the perspective of the physics of phonation, and setting aside for the moment the nasal cavity, the vocal 'system' is comprised of connected airways that allow an overpressure supplied by the lungs to produce an airflow at the mouth. This direct current

[2] In this thesis the term phonation is used to mean quasi-periodic sound production by the vocal folds and does not imply that the produced sound is intelligible. In the literature this may also be referred to as voicing.

(DC) flow from the lungs is modulated to produce an alternating current (AC) flow by the oscillations of the vocal folds. It is therefore convenient to divide the airway into two at the larynx, or more specifically at the glottis, i.e. the aperture between the vocal folds. Henceforth the supraglottal vocal tract between the glottis and lips is called the vocal tract, and the subglottal vocal tract including various branches of the lungs is called the subglottal tract (an illustration of these is shown in Figure 2-6).

2.2. Vocal folds

The vocal folds are a layered muscular structure located in the larynx. They are commonly described as comprising an inner 'body' and outer 'cover' layer, as shown in Figure 2-2. The body consists of the thyroarytenoid, or vocalis muscle, that increases in tension as it shortens, and a vocal ligament comprising the deep and intermediate layers of the lamina propria. These are covered by a mucous membrane made up of the superficial lamina propria and an epithelium. The importance of this layered structure to vocal fold motion will be discussed in the next section.

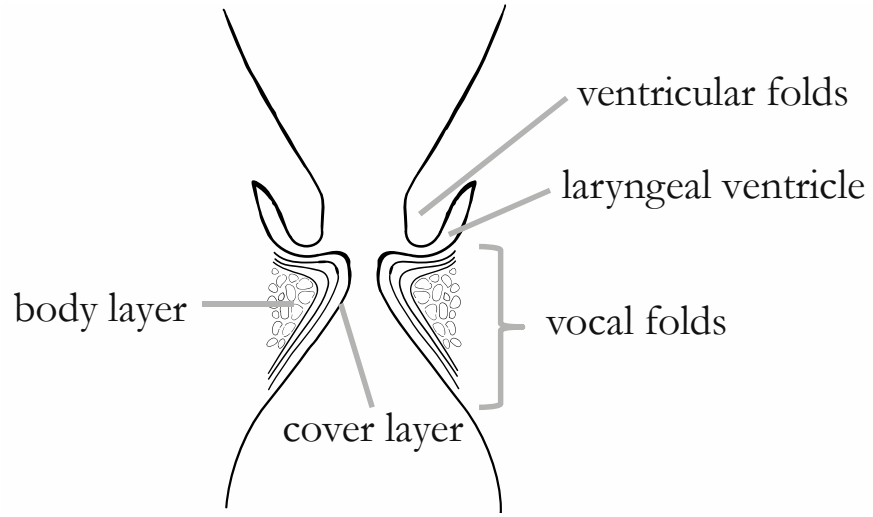


Figure 2-2 A simplified schematic of the vocal folds showing their layered structure and the position of the ventricular folds. Illustration by Olivia Cox.

The geometry of the glottis is described by whether the vocal folds are abducted (separated), or adducted (brought together) to form an open or closed glottis respectively. There are differences in the way this is achieved for phonation and pre-phonation. The limit of pre-phonatory vocal fold abduction may be achieved during inhalation, which yields the largest glottal aperture. These geometries are illustrated in Figure 2-3.

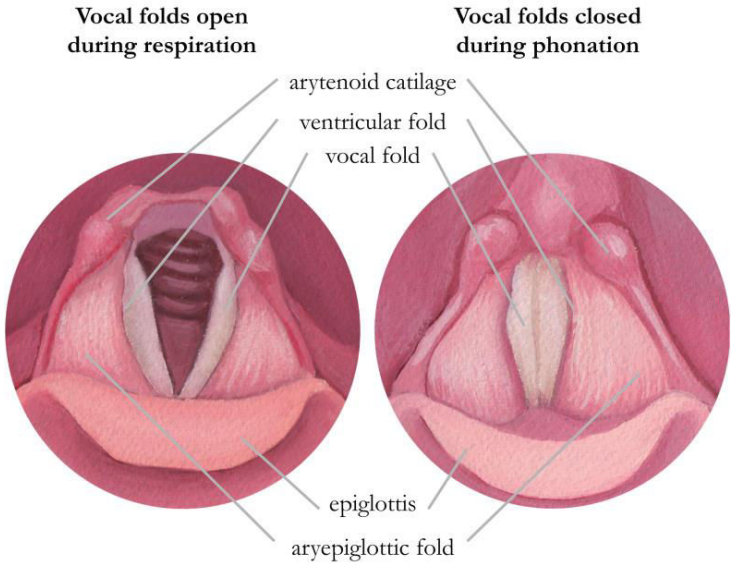


Figure 2-3 Simplified illustrations showing typical positions of the larynx during respiration and phonation. Illustration by Olivia Cox.

During phonation, the glottis alternates between closed or nearly closed, and open to a small fraction of the inhalation value. In phonation, the alternation is usually almost periodic and its fundamental frequency, f_0, is typically 80-200 Hz for men's speech, and 150-300 Hz for women's (Clark *et al.*, 2007). However, the upper limit of f_0 can exceed 1 kHz in singing.

The vibration of the vocal folds is not due to rapid muscular twitching but to the fluid-structure interaction between the airflow through the glottis and the moving walls. Muscular control is partially responsible for controlling the vocal fold geometry and tension, although the geometry also depends on aerodynamic factors. The vocal folds are attached to the thyroid and arytenoid cartilages, and so can be stretched by using the cricothyroid muscle to change the angle between the two cartilages, this type of movement is called cricothyroid tilt. Changes in tension can also be provided by the vocalis muscle, or by the lateral and transverse cricothyroid muscles.

The inferior (i.e. lower) parts of the vocal folds are connected directly to the wall of the trachea but their superior edge is linked to a somewhat similar structure called the ventricular folds, or the false vocal folds (due to their vibration in certain types of phonation). These two structures are separated by the Morgagni (or laryngeal) ventricles, see Figure 2-2. The superior portion of the ventricular folds is connected to the

aryepiglottic folds, which join the arytenoid to the epiglottis, and may also contribute to vibration in certain types of phonation. Collectively, the section of the vocal tract between the vocal folds and the epiglottis will be referred to as the aryepiglottic larynx[3].

2.2.1.Glottal cycle

To produce phonation it is necessary for the lungs to supply an excess air pressure below the glottis, the subglottal air pressure P_{sg}. A standard description based on Van den Berg's myoelastic-aerodynamic theory of voice production is given by e.g. (Clark *et al.*, 2007; Titze, 1994). For a glottal cycle that begins with the vocal folds fully adducted, P_{sg} must increase until it is sufficient to force the vocal folds apart and allow air to flow through the glottis. If the airflow through the glottis is laminar and lossless, then Bernoulli's law applies, so that at any point in the airway

$$(2\text{-}1)\quad P + \frac{1}{2}\rho v^2 = constant$$

where P is the pressure, ρ the density of air, and v the particle velocity that gives a flow Φ through cross-sectional area X of $\Phi = Xv$. Therefore, when the X is small (as is the case for a narrow glottis), a relatively high v must cause a drop in the pressure in the region between the vocal folds so that a so-called 'Bernoulli force' acts to adduct them. The closure of the vocal folds may allow P_{sg} to increase until it is sufficient to reopen the folds and to restart the cycle anew. A simplified schematic of the vocal fold motion during a single cycle is shown in Figure 2-4.

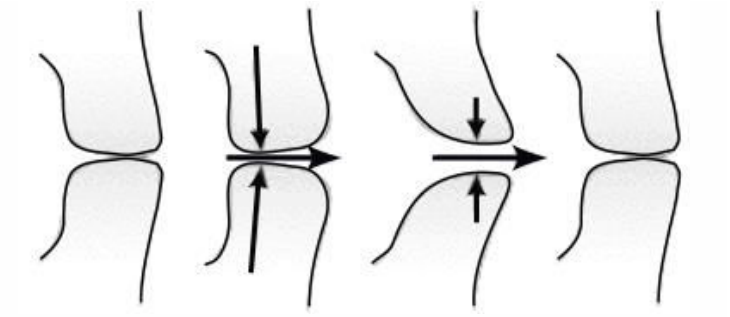


Figure 2-4. Side view of the vocal folds during a single cycle of oscillation. Reproduced with permission from Wolfe *et al.* (2009a). Airflow is indicated by horizontal arrows. The Bernoulli force is indicated by vertical arrows.

[3] In the literature this is variously referred to as the epilaryx, epilaryngeal tube, aryepiglottis, etc.

Note that the assumption of fully adducted vocal folds is not necessary and oscillation is possible even when the vocal folds do not completely close at any point in the cycle. After the initial transient phase, the oscillation of the vocal folds during phonation is principally periodic with fundamental frequency f_0.

Any turbulence generated due to separation of the air jet at the exit of the glottis means that the kinetic energy of flow through the glottis is lost, so that equation (2-1) no longer holds. In this case a non-zero average transglottal pressure $\Delta P = P_{sg}\text{-}P_{io}$ is produced, where P_{io} is the downstream or intra-oral pressure.

2.2.2.Models of vocal fold oscillation

An early attempt to quantify the motion of the vocal folds was made by Helmholtz (1875), who suggested representing the vocal folds motion as an outward striking valve in a one-dimensional waveguide, where pressure upstream (P_{sg}) would open the valve, and pressure downstream (P_{io}) would close the valve. Fletcher (1993) theoretically explored the conditions under which such a model would achieve self-oscillation, and found that an inertive load upstream (i.e. the subglottal tract) and a compliant load downstream (i.e. the vocal tract) were necessary. This was verified experimentally with a simple brass cantilever valve in a cylinder driven by air pressure (Tarnopolsky *et al.*, 2000). An alternative mode of oscillation, discussed by both Helmholtz and Fletcher, is the 'sliding door' model, in which pressure upstream and downstream can act to open the vocal folds.

An alternative representation of the vocal folds considers each of them as a single mass, m, on a spring with constant k (Flanagan, 1968). When there are no (or completely elastic) collisions the motion of each mass could be approximated as that of a simple harmonic oscillator, with a characteristic mechanical resonance frequency $f_m = 1/2\pi(k/m)^{1/2}$. The next level of complexity is achieved in models with two masses representing each fold, so that the time delay between the masses is considered e.g. (Ishizaka and Flanagan, 1972). This allows for consideration of the vocal fold geometry as convergent (upper part of the vocal folds closer together) or divergent (lower part closer together). Titze (2008) assumed that this model is appropriate to describe phonation, and used it to argue that an inertive vocal tract is always beneficial to driven vocal fold oscillation.

It is possible to consider any number of masses, e.g. Titze's 16 mass model (1974), and the n-mass model of Kob (2002). Furthermore, finite element models of the vocal folds have been suggested e.g. (Alipour *et al.*, 2000), in which more realistic geometry and vocal fold

properties are modelled by a large number of individual elements. Such models require many input parameters including the visco-elastic properties of all relevant parts of the vocal folds, which are not readily available, so many assumptions are necessary. Furthermore, accurate modelling of collisions, turbulence and changing boundary conditions are still beyond current capabilities.

These mass-spring vocal fold models contain many simplifications to the fluid-structure interaction. Empirical data can be obtained from *in vitro* self-oscillating replicas of the vocal folds to provide information about the aerodynamics in addition to the acoustics e.g. (Lucero *et al.*, 2009; Ruty *et al.*, 2007; Van Hirtum *et al.*, 2007).

2.2.3.Mechanical resonances of the vocal folds

The nonlinearity of the Bernoulli term and inelastic vocal fold collisions produce many higher harmonics at multiples $2f_0$, $3f_0$ etc. Further, due to the layered structure of the vocal folds and the three-dimensional nature of their movement during a cycle, there is no single f_m and therefore no simple relationship between it and f_0. The only known *in vivo* study of the mechanical response of the vocal folds by mechanical excitation at the neck reports at least three mechanical resonance peaks in the range 100-350 Hz for f_0 = 110 Hz (Švec *et al.*, 2000). Response spectra are shown in Figure 2-5.

The structure of the vocal folds alone does not directly determine f_0. Nonetheless, the layered vocal fold structure may provide a qualitative explanation for other aspects of vocal fold motion: the mechanisms, or registers, of the voice. Titze (1994) explains that, when the vocal muscle tension is the dominant factor in determining f_0, the outer layers vibrate passively, so the relatively large combined mass in motion and strong collisions produce an oscillation with rich harmonic content known as Mechanism 1 (M1), modal or chest voice register. Mechanism 2 (M2), also known as falsetto or head voice, is produced when the vocal fold is lengthened and so the main body layer, the vocalis muscle, remains stationary, so that the frequency is determined by the movement of the ligament and mucosal surface waves. Collisions therefore involve less change in momentum or do not occur (*cf.* Ferrein's bowed string analogy (Ferrein, 1741)), yielding weaker high harmonics. This difference is the chief distinction between M1 and M2.

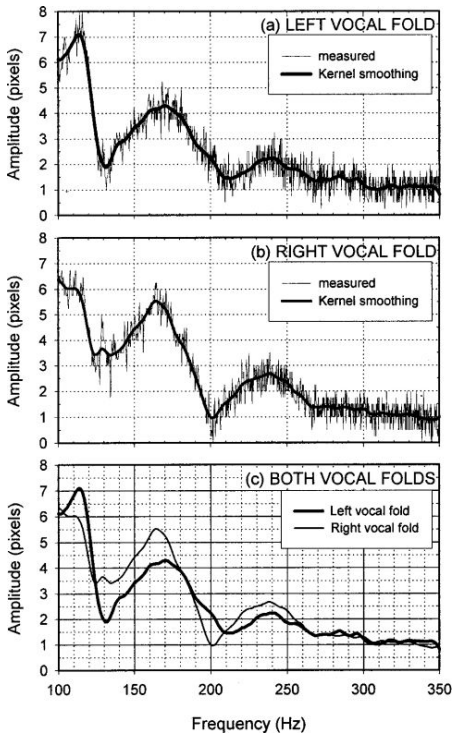


Figure 2-5. Example mechanical response reproduced with permission from Švec *et al.* (2000), showing a) the left vocal fold, b) the right vocal fold, and c) smoothed curves for both vocal folds. Three mechanical resonances peaks appear at ~110, 170 and 240 Hz.

Aspects of the vocal fold motion can be monitored indirectly with electroglottography (EGG), in which skin electrodes are used to pass a small, high-frequency electrical current through the neck, at about the level of the glottis. The changing electrical admittance due to the varying vocal fold contact provides a signal that is assumed to be a monotonic function of the glottal contact. The transition from M1 to M2 is apparent from changes to the shape of the EGG signal. The length of the closed phase is reduced, and the amplitude is reduced, signifying reduced contact between the vocal folds, and a rapid increase in f_0 (Roubeau *et al.*, 2009).

f_0 is important in tonal languages where it provides lexical contrast but even in non-tonal languages it performs functions of stress and intonation (Gussenhoven, 2004, pp. 12–25). f_0 is determined by a combination of biomechanical parameters i.e. motion of vocal fold layers and their muscular control; and aerodynamic parameters P_{sg} and $\boldsymbol{\Phi}$.

2.2.4.Pressure-frequency relationship

The P_{sg}, f_0 relationship has been studied from a theoretical perspective (Titze, 1989) and the maximum physiological rate of f_0 change may provide articulatory constraints that influence the production and shape of tone contours in speech (Xu and Sun, 2002). It is thus important to quantify how much f_0 change depends on mechanical and aerodynamic parameters.

There are few experimental data on the *in vivo* P_{sg}, f_0 relationship, since direct measurement of subglottal pressure requires a very invasive tracheal puncture (Plant and Younger, 2000; Sundberg *et al.*, 2013), and living subjects may change mechanical and laryngeal parameters unconsciously and automatically (Baer, 1979). For these reasons many investigations on excised larynges have been performed following the work of Ferrein (1741), and Van den Berg and Tan (1959) (a review of more recent work is given by Alipour and Scherer (2007)). Such *ex vivo* studies have the advantages of controlling laryngeal parameters independently, thus avoiding automatic compensation effects known to occur *in vivo*.

The dependence of f_0 on P_{sg} is widely reported as the local gradient (df/dP), with typical increases of 0.1 to 0.7 Hz for an increase in pressure of 1 kPa or (1-7 Hz/cm H_2O (Baer, 1979; Titze, 1989; Van den Berg and Tan, 1959) but values as high as 2.0 Hz/kPa can be found (Alipour and Scherer, 2007; Lieberman *et al.*, 1969). In practice, the P_{sg}, f_0 relationship is nonlinear. It also exhibits dependence on the degree of vocal fold elongation and adduction (Alipour and Scherer, 2007; Hsiao *et al.*, 2001), and varies significantly between living human subjects (Plant and Younger, 2000).

Discontinuities in the oscillation of the vocal folds have also been reported for excised human larynges. Pitch jumps associated with changes of laryngeal mechanism or transition to a subharmonic oscillation may occur as a consequence of minor changes in vocal fold tension (Švec *et al.*, 1999) or crico-thyroid tilt (Horáček *et al.*, 2004).

For practical reasons, excised larynx experiments are often performed on species other than humans, e.g. (Berry *et al.*, 1996; Jiang and Titze, 1993). However, although non-human larynges may be used for assessing aerodynamic and acoustical models, they can differ substantially from those of humans in biomechanical properties of the vocal folds, laryngeal anatomy, phonation threshold pressure (Mau *et al.*, 2011), and P_{sg}, f_0 behaviour (Alipour and Jaiswal, 2008). An inter-species comparison of the P_{sg}, f_0 relationship in excised larynges has recently been reported by Alipour *et al.* (2013).

In addition to differences in the P_{sg}, f_0 relationship, studies of excised canine, porcine, ovine and bovine larynges have noted that the supra-glottal tissues and structures in some animal larynges can cause a non-zero supraglottal pressure, which may have an effect on the glottal resistance (Alipour *et al.*, 2007; Alipour and Jaiswal, 2009). Although the same observation is likely true for the aryepiglottic larynx in humans, the geometry of the supraglottal structures is not directly comparable inter-species, and the control of these structures *in vivo* is not well understood. *In vivo* measurements and numerical models have shown that the ventricular folds can enhance, or suppress vocal fold vibrations (Bailly *et al.*, 2014).

Both P_{sg} and f_0 have been shown to be correlated with sound pressure level *in vivo*, with increases of 10 dB related to: a five semitone increase in pitch (Gramming *et al.*, 1988), and to a doubling of P_{sg} (Sundberg *et al.*, 1993).

In summary, properties of the voice source can be determined through *ex vivo* experimentation. The advantages of these experiments are the ability to control and monitor the laryngeal parameters, precisely and independently, which may be difficult to access and/or impossible to operate independently *in vivo*, and to avoid unconscious compensation effects. The P_{sg}, f_0 relationship provides a point of comparison between aerodynamic and mechanical methods of vocal fold vibration control. However, there remain open questions about its global form, which is non-linear, and its dynamic non-linear behaviour, which is not well understood. Further the role of the aryepiglottic larynx in normal phonation is not well understood. These questions are the focus of Chapter 5.

2.3. The Vocal Tract

The vocal tract is the airway between the lips and the vocal folds, including the oral cavity, pharyngeal cavity, and the epilaryngeal tube, as shown in Figure 2-6. The length of the vocal tract is typically considered as 17-20 cm for adult males and 14.5-17 cm for adult females (Sundberg, 1987). The vocal tract length determined from magnetic resonance imaging (MRI) data is typically shorter, ~14-16 cm and ~13-15 cm for adult men and women respectively (Fitch and Giedd, 1999), although some of this difference may be accounted for by an increase of order 1 cm in larynx height when subjects are in the supine position for MRI measurements (Traser *et al.*, 2013).

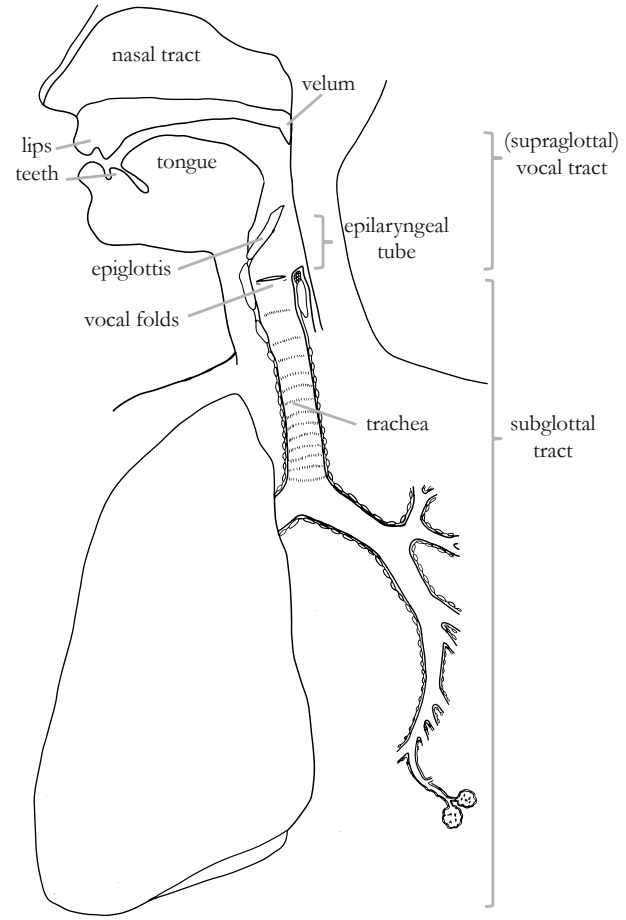


Figure 2-6 A simplified sketch showing the nasal tract, vocal tract and subglottal vocal tract. The aryepiglottic larynx between the vocal folds and the epiglottis is labelled epilaryngeal tube. Illustration by Olivia Cox.

The aryepiglottic larynx is the short section of the vocal tract between the vocal folds and the epiglottis, where the ventricular folds are located. It is separated from the posterior wall of the pharynx, which leads to the oesophagus, and may be closed completely, as in swallowing, by the epiglottis.

Note that the nasal tract from the nares (nostrils) to the soft palate (velum) may be coupled in parallel to the vocal tract by opening the velum, to produce nasal consonants and nasalised vowels. However, in all vocal tract configurations presented in this thesis the velum is assumed to be closed, as is the case for non-nasalised vowels, and thus the nasal tract is not considered as part of the vocal tract. The subglottal tract is the continuation of the respiratory system below the glottis, and is discussed in 2.4.

The primary roles of the vocal tract in phonation are to transmit DC air flow and acoustic waves and to act as an acoustic filter for the harmonic signal produced by the vocal folds[4]. The vocal tract is effectively an acoustic duct that can be adjusted in length and cross-sectional area by movement of the articulators; jaw, lips, teeth, tongue, velum, pharynx, epiglottis, aryepiglottic folds, ventricular folds, and vocal folds to alter the acoustic resonances, i.e. the gain spectrum of the filter. The consonants of the International Phonetic Alphabet can be categorised in terms of the placement of the vocal tract articulators, and adjusting the articulators during sustained phonation allows humans to produce a wide range of vowel sounds and lateral and rhotic consonants (Clark *et al.*, 2007; Wolfe *et al.*, 2009b), as shown in the example MRI images in Figure 2-7 (Valdes, 2013). The time variation in the filter function is also important in many consonants.

Estimates of the vocal tract geometry have been obtained from X-ray (Fant, 1960; Johansson *et al.*, 1985) and MRI studies (Baer *et al.*, 1991; Demolin *et al.*, 1996; Story *et al.*, 1996). With such information it is possible to compute a three-dimensional finite element wave solution to the acoustic waves in the vocal tract. However, this approach relies on the relatively poor precision of the geometric data. Therefore a one-dimensional waveguide is a reasonable simplification with the advantage that the underlying physics is highlighted.

[4] The vocal tract can act as both the filter and the source of sound by constricting the airflow with various articulators (glottis, teeth, tongue), causing turbulence and producing the broadband signal characteristic of fricatives, e.g. [h, s, f]. This function of the vocal tract is not considered directly in this thesis.

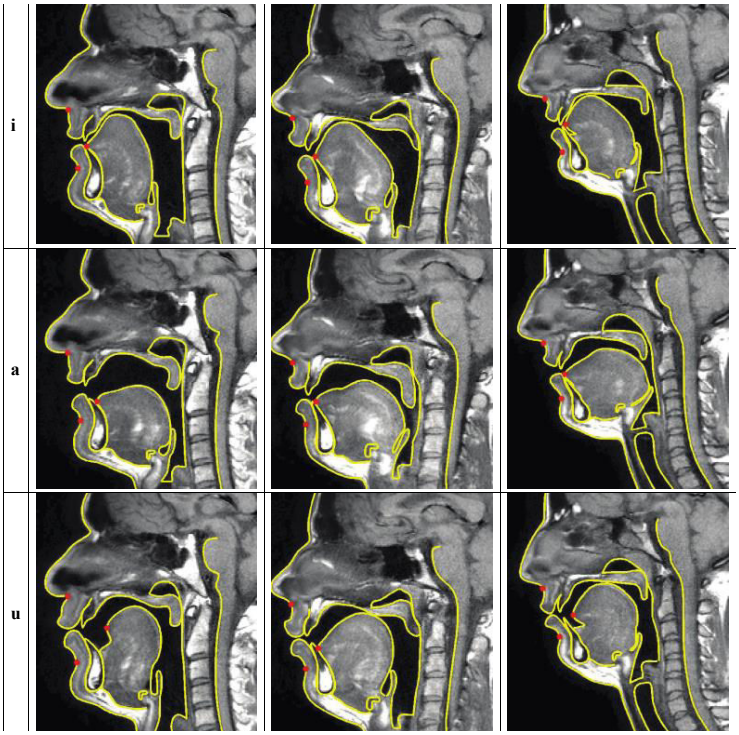


Figure 2-7 MRI scans showing the French vowels [i, a, u] pronounced by male speakers (left, middle), and a female speaker (right). The contours of the vocal tract and fixed reference points are highlighted in yellow. Reproduced with permission from Valdes (2013, p. 114).

2.3.1.Modelling the vocal tract

For frequencies below a few kilohertz, wavelengths are much greater than the transverse dimensions of the vocal tract, so a simple one-dimensional approximation of the vocal tract as a rigid cylinder captures much of the physics. Curvature of such a tube has little effect in the frequency range studied (Kob, 2002; Sondhi, 1986), so can usually be neglected. At higher frequencies, deviations from cylindrical geometry become more important, and at frequencies below a few hundred hertz, the non-rigidity of the walls needs to be considered (Makarov and Sorokin, 2004; Sondhi, 1974). For most types of phonation, the glottal aperture is much smaller than the cross-sectional area. For such cases, the vocal tract can be considered as a pipe with a closed end at the glottis. A highly simplified straight vocal tract is shown in Figure 2-8

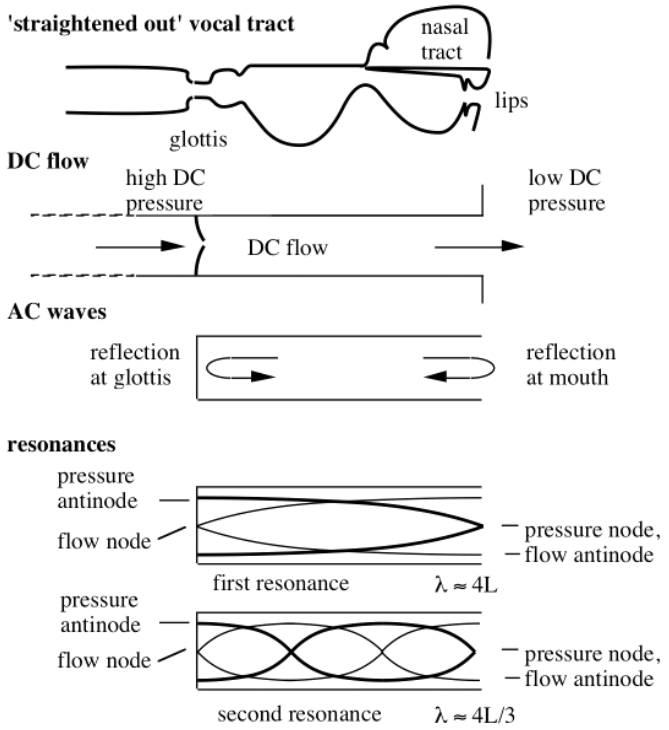


Figure 2-8 Simplified vocal tract model showing DC flow, AC waves and pressure and flow resonances. Reproduced with permission from Wolfe *et al.* (2009a).

The acoustic impedance is the ratio of sound pressure (the AC component of pressure) to volume flow (Φ, measured in $m^3 \cdot s^{-1}$). The acoustic impedance is a complex quantity. The real term represents a resistive part where pressure and flow are in phase. Positive resistance converts mechanical energy into heat. The imaginary part represents energy storage; it can be either positive or negative, so that the impedance can be inertive (pressure leads flow) or compliant (pressure lags flow). For a plane wave propagating in a cylinder terminated at the remote end by a load impedance Z_L, the impedance at the input end is given by the following relation (Fletcher and Rossing, 1998).

$$(2\text{-}2) \quad Z = \frac{\rho c}{\pi r^2}\left(\frac{\pi r^2 \cdot Z_L \cos(\Gamma l) + 2j\rho c \cdot \sin(\Gamma l)}{j\pi r^2 \cdot Z_L \sin(\Gamma l) + 2\rho c \cdot \cos(\Gamma l)}\right)$$

where ρ is the density of air, c is the velocity of sound, r the radius, l is the length and Γ is the complex wave number

$$(2\text{-}3) \quad \Gamma = \frac{\omega}{c} - j\alpha \approx \frac{\omega}{c} - j\frac{1.2\times10^{-5}\sqrt{\omega}}{r}$$

where ω is the angular frequency and α is an attenuation coefficient, which accounts for losses at the tube walls due to viscous drag and thermal conduction. In general, the impedance of one end of a duct does not completely specify the system, information of its geometry is also required. However, the assumption of a cylindrical model of the vocal tract is sufficient to account for the neutral vowel geometry considered in Chapters 3 and 4.

The formants (peaks in the output sound spectrum) of the neutral vowel [ə:] occur with approximately even spacing. This behaviour can be replicated by a cylinder ideally closed at the far (glottis) end, i.e. $Z_L = \infty$

$$Z_{\text{closed}} = -j\frac{\rho c}{\pi r^2}\cot(\Gamma l) \quad (2\text{-}4)$$

In this case the impedance minima, i.e. resonances *R*i that give rise to formants, occur at $\lambda_{Ri} = 4l, 4l/3, 4l/5$ etc., and impedance maxima, or anti-resonances *A*i, occur at $\lambda_{Ai} = 2l, 2l/2, 2l/3$ etc., where i is an index that increases with frequency.

The resonances act as an acoustic impedance matcher from the high impedance of the glottis to the low impedance of the radiation field at the open mouth. Thus the resonances produce peaks in the spectral envelope of the sound radiated from the mouth; the formants *F*i (Fant, 1960).

Various terms describe vocal tract resonances, formants and vocal fold frequencies in the literature. This thesis follows the terminology of Wolfe (Wolfe, n.d.): The acoustic vocal tract resonances *R*1, *R*2, *R*3 produce formants *F*1, *F*2, *F*3 etc., and the fundamental frequency of the vocal folds is f_0, with harmonics $2f_0$, $3f_0$ etc. Furthermore, the properties of the resonances and formants are denoted such that the frequency of *R*1 is f_{R1} and its bandwidth is B_{R1}.

For a closed-open cylinder, the spacing of f_{Ri} is comparable with the regular spacing of f_{Fi} in the neutral vowel [ə:]. For a typical male vocal tract length of 17 cm the vowel has formants with f_{Fi} ~ 500, 1500, 2500 Hz etc.

For other vowels, this simple cylindrical model is not appropriate. In such cases, the vocal tract can be modelled as a series of cylinders of different *l* and *r*. To calculate the impedance iteratively, each cylinder starting at the glottis end provides the Z_L of the following cylinder in equation (2-2). This allows the use of more realistic geometries,

which can be approximated for example from MRI imaging, and used to determine the area function of the tract, e.g. (Story *et al.*, 1996). Birkholz (2005; 2007) produced a virtual vocal tract for voice synthesis with adjustable realistic geometry, as shown in Figure 2-9.

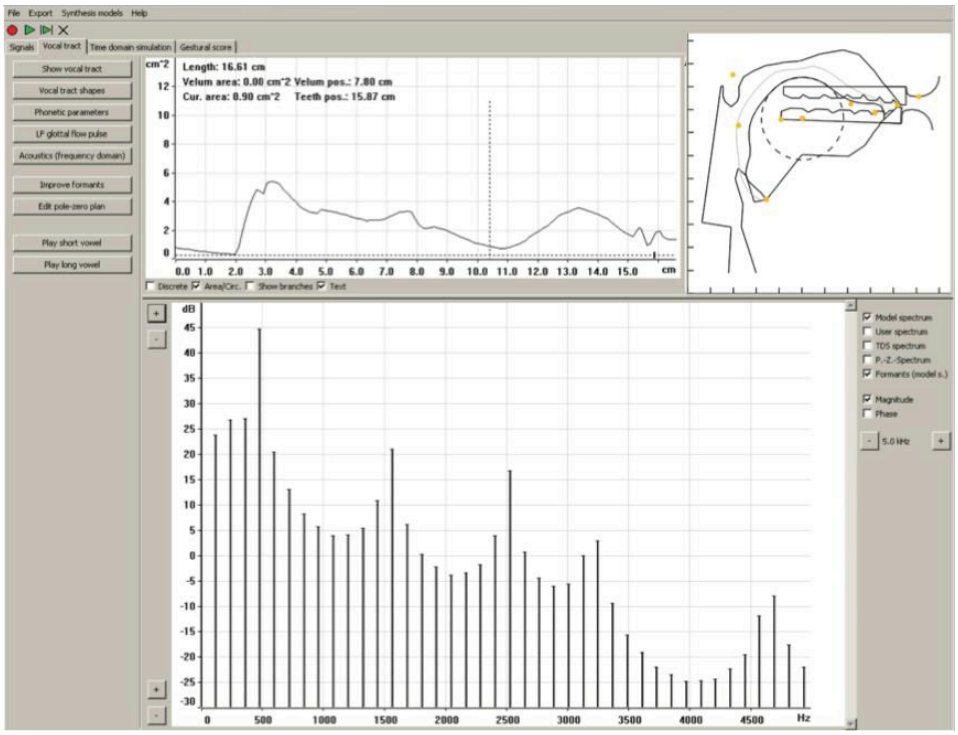


Figure 2-9 Example vocal tract area function (top left) with corresponding simplified vocal tract geometry (top right) and harmonic spectrum outside the mouth (bottom) for a neutral vowel. Figure adapted from a screen capture from the Vocal Tract Lab software of Birkholz (2005; 2007).

2.3.2.Importance of vocal tract filter

Formants are important for speech perception. The frequency and relative amplitude of *F*1 and *F*2 and sometimes *F*3 largely characterise the vowel sounds of non-tonal languages, such as English. A plot of the f_{F2}, f_{F1} plane maps the range of vowel sounds in a language. The axes are traditionally plotted backwards to correspond to the axes used by phoneticians and linguists: jaw height vs the position (front to back) of the tongue constriction. Increasing jaw height gives increasing lip aperture and increasing f_{F1}. Front tongue constriction gives high f_{F2} and *vice versa*. Such a plot is known as the vowel plane. A vowel plane of the perceived vowels in Australian English is shown in Figure 2-10.

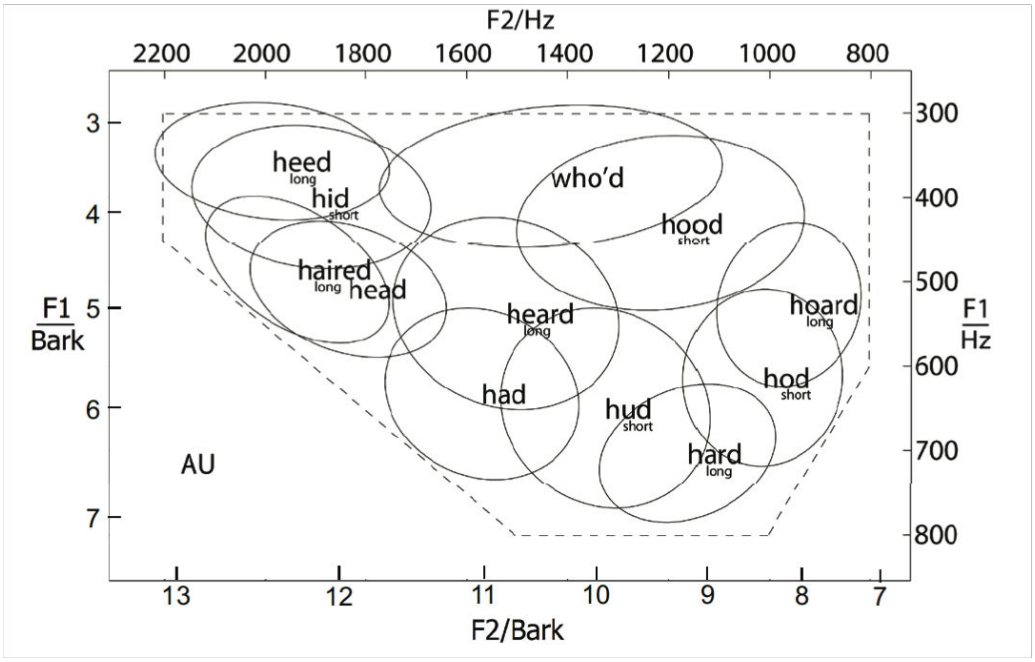


Figure 2-10 Perceptual vowel plane for Australian English, determined from an automated study of perceived vowels, reproduced with permission from Ghonim *et al.* (2013). The centres of the ellipses show the average f_{F1} and f_{F2} of sounds that listeners perceive as the labelled vowel sounds. The semi-axes show the standard deviation in the direction of correlation and in the perpendicular direction. The h(vowel)d context is used because almost all such combinations are real English words.

Although the formant centre frequencies determine the vowel plane in Figure 2-10, the filtering effect applied to the source signal also depends on the amplitudes and bandwidths of the formants, and thus on those of the vocal tract resonances, all of which are perceptually important. The bandwidth B of the formants provides valuable information for listeners (Brown *et al.*, 2010; Summerfield *et al.*, 1985), for speaker identification (Childers and Diaz, 2000; Childers and Wu, 1991; Hanson and Chuang, 1999), and about the glottal source (Childers and Wu, 1991; Ishizaka and Flanagan, 1972; Rabiner and Schafer, 1978; Rothenberg, 1981). B is often normalised, inverted and expressed as the quality factor $Q = f/B$. Wide B and therefore low Q resonances would provide insufficient boost to harmonics to be recognizable as distinct vowel sounds, whereas narrow B resonances increase intelligibility. However, losses in the human vocal tract place an upper limit on the magnitude of Q, and consequently synthesised voices with high Q resonances sound unnatural.

Tuning a resonance for musical purposes is an interesting extension of vocal tract filtering. For example, adjusting $R1$ to match f_0 (and thus changing the vowel) is a technique used by

sopranos to sing the upper octave of the standard soprano register from *C5* to *C6* (Garnier *et al.*, 2010; Henrich *et al.*, 2011; Joliveau *et al.*, 2004a, 2004b; J. Sundberg, 1973), or other singing styles in which a harmonic of f_0 coincides with a vocal tract resonance (Bourne and Garnier, 2012; Henrich *et al.*, 2011, 2007). These techniques are used to efficiently transfer energy in order to boost the loudness of the voice, at the expense of vowel intelligibility.

2.3.3. Measuring the acoustic properties of the vocal tract

Accurately measuring the most useful acoustic properties of the vocal tract is difficult, chiefly due to the lack of access to the glottis. For this reason, data on the amplitudes and bandwidths of formants and resonances are scarce in the literature. However, the data that do exist display the ingenuity of voice researchers in overcoming the tyranny of distance.

Measurements of the formants in the radiated voice sound give an estimation of *B* with a resolution comparable with the fundamental frequency f_0, e.g. (Bogert, 1953; Peterson and Barney, 1952). Improved frequency resolution has been achieved with swept sine excitation at the neck with the glottis held closed (Dunn, 1961; Fant, 1972; Fujimura and Lindqvist, 1971; Van den Berg, 1955) and held open (Fujimura and Lindqvist, 1971). Measurements during phonation have also been demonstrated using a broadband vibration applied outside the neck (Pham Thi Ngoc and Badin, 1994). However, these measurements are subject to unknown transfer functions at the neck, so that the actual frequency response of the sound injected in the vocal tract is unknown. Attempts to provide an impulsive signal by popping a balloon inside the tract could overcome this issue but have low signal:noise ratio and have not yielded repeatable results (Coffin, 1989; Russell and Cotton, 1959). Similarly, a microphone inserted via the nasal cavity to record the sound close to the glottis is uncomfortable and the frequency response is sensitive to the position. A preliminary study to determine the vocal tract transfer function in this way did not yield an improved spectral response (Kob, 2002). Measurements with external microphones and broadband excitation at the open mouth with an electric spark (House and Stevens, 1958; Tarnóczy, 1962); a starting pistol, by the speech researchers at KTH Stockholm, which appears not to have been reported formally (Ternström via Wolfe: personal communication); or a synthesised broadband signal (Epps *et al.*, 1997) are made in parallel with the radiation impedance of the open mouth, so that the magnitude of the impedance is not accurately known and thus it is impractical to calculate *B*.

The methods of Pham Thi Ngoc and Badin (1994) and Epps *et al.* (1997) allow high precision measurements during phonation, and these demonstrate that there are differences

in *B* between gestures with a closed glottis and those with phonation (Pham Thi Ngoc and Badin, 1994), and also shifts in the resonance frequencies f_{Ri} in whisper, creak and normal voice (Swerdlin *et al.*, 2010). Furthermore, attempts to reconstruct phonated speech from whispered speech (Lv and Zhao, 2009) have demonstrated that it is necessary to take into account not only the observed shift in formant frequencies (Jovicic, 1998; Swerdlin *et al.*, 2010) but also a change in *B*. However, the most commonly used bandwidth data for modelling the voice is based on closed glottis measurements (Childers and Diaz, 2000; Hawks and Miller, 1995), which therefore may not be representative of the voice during phonation.

Those observations led to the present study. In Chapters 3 and 4 measurements of the response of the vocal tract to broadband excitation at the lips are presented. This gives an impedance spectrum with precise measurements of the frequency, amplitude and bandwidth of the resonances both without and with phonation. It also avoids the disadvantages detailed above. The magnitudes and bandwidths of the resonances from this impedance spectrum give good estimates of the losses in the vocal tract, which could be useful in, for example, voice synthesis.

2.3.4.Non-rigid vocal tract walls.

The vocal tract is not rigid. Vibration of its walls is demonstrable and the distribution and amplitude around the face and neck under internal excitation have been mapped by Fant *et al.* (1976) as shown in Figure 2-11. The influence of the non-rigidity of the walls at low frequency is likely to be important for the characteristic broadband sound in the production of plosives (Delebecque *et al.*, 2013; Kewley-Port, 1983; Stevens, 2000). Resonance frequencies determined using area functions derived from MRI and calculated using a rigid walled assumption differ from experimentally determined formants, particularly at low frequency (Makarov and Sorokin, 2004).

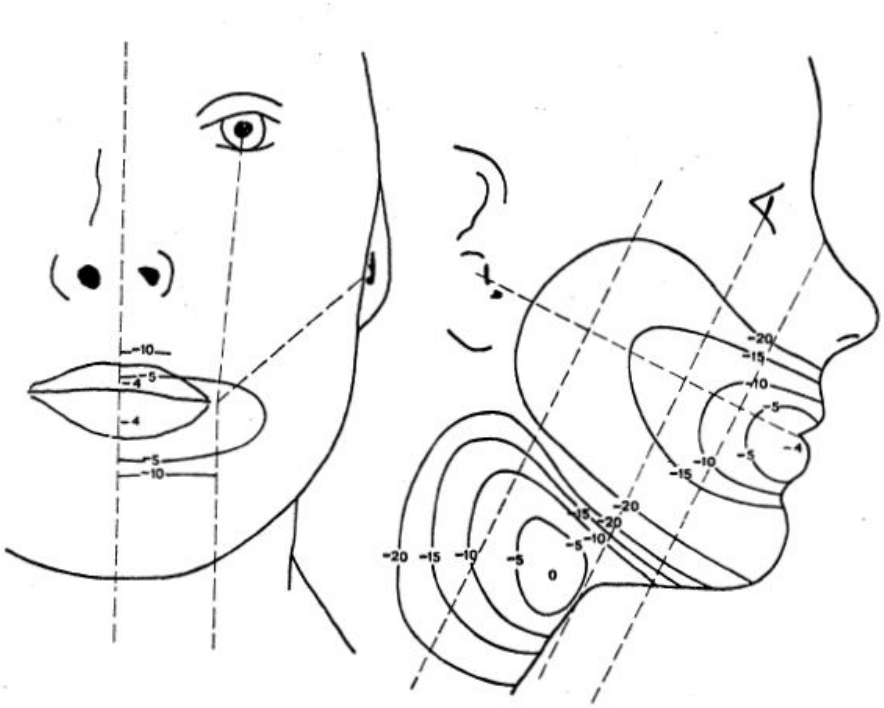


Figure 2-11 Contours of equal vibrational amplitude along the face and neck measured externally with an accelerometer, while low frequency sound is injected from a narrow tube through the lips. Reproduced with permission (Fant *et al.*, 1976).

Models for this low frequency mechanical behaviour of the vocal tract have been proposed to describe the expected mechanical resonances of the tract (e.g. (Fant, 1972; Ishizaka *et al.*, 1975; Makarov and Sorokin, 2004; Sondhi, 1974). These models apply electrical transmission line theory, with acoustic pressure represented by electrical potential and volume flow by current. In such models acoustic inertance is analogous to electrical inductance, and acoustic compliance to electrical capacitance. This transmission line approach is appropriate when the wavelength is large compared to the dimensions of the vocal tract, then one can use a series of simple lumped elements for components that are small in comparison to the wavelength of interest. However, the models lack experimentally measured parameters.

Summaries of the available data (Fant, 1985; Stevens, 2000; Wakita and Fant, 1978) generally refer to two studies: Fant *et al.* (1976) excited the tract acoustically and found broad mechano-acoustic resonances around 200 Hz, similar to the lowest f_{F1} reported with swept sine excitation (Fujimura and Lindqvist, 1971) and under increased ambient air pressure conditions (Fant and Sonesson, 1964). Ishizaka *et al.* (1975) used mechanical

vibration to excite the cheek and neck and found broad mechanical resonances in the range 30-70 Hz. Each of these studies considered only a single mechanical resonance in isolation.

It appears that there has been no recent work to improve the estimates of parameters from these studies. The values determined in the Ishizaka *et al.* study are used to model non-rigid walls in the vocal tract geometry based voice synthesis models 'TubeTalker' (Story, 2011, 1995) and 'Vocal Tract Lab' (Birkholz, 2012, 2005; Birkholz *et al.*, 2007).

In Chapter 3, an acoustic impedance head is used to measure the impedance of the vocal tract, which measures the acoustic and mechanical resonances and anti-resonances simultaneously. The measurements are extended to low frequency to give new estimates of the yielding wall parameters. Using a simple model, these measurements relate the impedance magnitudes, frequencies and bandwidths to tissue mechanical parameters, which are necessary for accurate modelling of the low frequency response of the vocal tract.

2.4. The Subglottal Vocal Tract

The subglottal tract begins below the glottis from where the trachea continues for approximately 10 cm before branching into two parts. Subsequent branches continue asymmetrically and are eventually terminated by the many bronchioles of the lungs. A simplified representation of the branching is shown in Figure 2-6. In contrast to the vocal tract, the geometry of the subglottal tract is relatively fixed: there is little possibility of adjusting the position or shape, except presumably to make changes by expanding the walls under high pressure, and by changing the volume of the terminating air sacs. The subglottal tract is primarily of importance for connecting the lungs to the glottis and thus allowing a positive average pressure P_{sg}, which is controlled predominantly by the muscles of the torso. However, it is also a resonant duct that provides an acoustic load on the upstream side of the vocal folds.

To first approximation it may appear that there are no strong resonances and the subglottal tract acts like a pure resistance, particularly when viewed through the small aperture, and hence, high impedance, of the glottis, as discussed in chapter 4. However, direct and indirect measurements of the subglottal system have identified several resonances below a few thousand hertz although there is not strong agreement on the frequencies at which they occur.

2.4.1.Importance of the subglottal tract

The influence of the subglottal tract may be seen in the pressure transfer function measured at the mouth with certain types of phonation, when the glottal aperture is large (Fant *et al.*, 1972; Fujimura and Lindqvist, 1971). Although the distinctive features of the subglottal tract transfer function do not appear to have been fully explored, more recently accelerometers placed on the neck have been used to monitor the subglottal 'formants' during speech in order to discern their effect on the voice e.g. (Chi and Sonderegger, 2007, 2004; Lamarche and Ternström, 2008; Lulich, 2010; Zañartu *et al.*, 2013). It is assumed that the resonances of the subglottal tract are relatively fixed, so they may provide useful information for speaker normalisation in analysis of speech signals, i.e. knowing who is speaking (Wang *et al.*, 2008). Zhang *et al.* (2006) used *in vitro* experiments to show that subglottal resonances can cause discontinuites in the behaviour of a vocal fold in a model system, suggesting that it might also have acoustic effects *in vivo*. It has been suggested that such an effect may cause the global division of the vowel plane into front and back vowels that occur either side of the second subglottal resonance (Chi and Sonderegger, 2004; Lulich, 2010, 2006). Finally, the acoustic impedance of the subglottal tract at very low frequencies during respiration has been explored for its potential clinical relevance to detecting lung pathologies, e.g. (Brown *et al.*, 2010; Campbell and Brown, 1963; Diba *et al.*, 2011; Fredberg and Hoenig, 1978; Habib *et al.*, 1994; Harper *et al.*, 2001; Hudde and Slatky, 1989; Robinson *et al.*, 2011).

2.4.2.Measurements of the subglottal tract

Data on the morphology of the subglottal tract come from three main sources: Van den Berg's dissections of humans and canines (van den Berg, 1960); Weibel's (1963) measurements of excised lungs and casts; and Horsfield *et al.* who made measurements on casts of the lungs (summarised in (Horsfield *et al.*, 1971)).

As for the acoustic properties of the subglottal tract the resonances are not directly associated with the individual branches of the tract. For a one-dimensional acoustic waveguide model, the resonances are due to reflection at changes in cross-sectional area, so it is the total area of each generation of branches that must be considered. Van den Berg (1960) made the first acoustic impedance measurements on cadaveric human and canine lungs using swept sine excitation. He determined the first resonance to be ~300 Hz, a value that has never been verified. Ishizaka *et al.* (1976) made the first *in vivo* acoustic impedance measurements on laryngectomised patients, and subsequently measured

subglottal resonance frequencies have been closer to these values, i.e. a first resonance close to 600 Hz e.g. (Cranen and Boves, 1987; Lulich *et al.*, 2011; Sundberg *et al.*, 2013). For a full overview of published subglottal resonance and anti-resonance frequency data see Lulich (2006, pp. 41–43).

2.4.3.Models of the subglottal tract

Models of the subglottal tract have generally used frequency domain electrical analogues, like those discussed in 2.3.1 to combine acoustic data with geometry (typically from (Weibel, 1963) or (Horsfield *et al.*, 1971)) e.g. (Habib *et al.*, 1994; Harper *et al.*, 2001; Hudde and Slatky, 1989; Ishizaka *et al.*, 1976; Zañartu *et al.*, 2013). This has also been performed in the time domain model of subglottal tract (Ho *et al.*, 2011). Many simplifications are made in these models but comparisons of symmetric and asymmetric electrical analogues have shown that below 2 kHz the differences are small, and that the rigidity of the airway walls is an important factor in determining *B* (Fredberg and Hoenig, 1978), and modelling only a few generations is necessary, so that the termination impedance does not need to be considered (Lulich, 2006).

More recently Lulich *et al.* (2011) showed that based on data from accelerometer signals below the glottis, the subglottal tract is well approximated as a cylinder of acoustic length ~20 cm (related to the height of the subject, divided by an empirically derived scaling factor) for the purposes of predicting the second and third resonances of the tract. They state that this underestimates the first subglottal resonance frequency, and proposed a model including non-rigid walls to explain the difference.

2.4.4.Coupling the vocal tract and subglottal vocal tract via the glottis

The acoustic loading experienced by the vocal folds is likely to be a combination of the loads of the vocal tract and subglottal tract. In the swinging door models discussed in 4.2.2.2, the motion of the vocal folds is driven by the series impedance of the two impedances. However, in the sliding door model, the average pressure in the glottis would be the driving force, and it has been proposed (Stevens, 2000) that the driving force would be the addition of the two impedances in parallel. In Chapter 4, both of these models are compared.

2.5. The Source-Filter Model

2.5.1. Overview of linear source-filter model

The source-filter model (Fant, 1960) has enabled great progress in our understanding of voice production, particularly for voice analysis and synthesis. The simplest model treats the source (from the vocal folds) and the filter (the effect of the vocal tract) as independent. This would suggest that aerodynamic and acoustic effects due to the vocal tract are assumed to have no influence on the vibration of the vocal folds, and the details of vocal fold vibration do not affect the filtering. This approximation allows a qualitative and semi-quantitative understanding of voice production, particularly in cases when f_0 is lower than the frequency of the first acoustic resonance of the vocal tract, as is typically the case in speech.

Consider phonation with f_0 = 150 Hz: a vocal tract with resonances as described in section 2.3.1 would have the effect of boosting the transmission of the third harmonic $3f_0$ = 450 Hz and the 10th harmonic $10f_0$ = 1500 Hz, so that the envelope of the frequency spectrum measured outside the mouth would have peaks at these frequencies, as shown in Figure 2-12. The peaks in the spectral envelope (the formants *F*i) in this example are close to the resonances of the vocal tract. However, since the spectrum is sampled at f_0, the f_{Fi} do not provide precise information about vocal tract resonances, notably in the case of high female voices. Broadband noise due to airflow through the glottis provides some additional information at frequencies other than the harmonics of the voice. However, the frequency of the formants f_{Fi} has an expected error of at least 10% of f_0 when determined with standard methods of linear predictive coding (Monsen and Engebretson, 1983; Vallabha and Tuller, 2002).

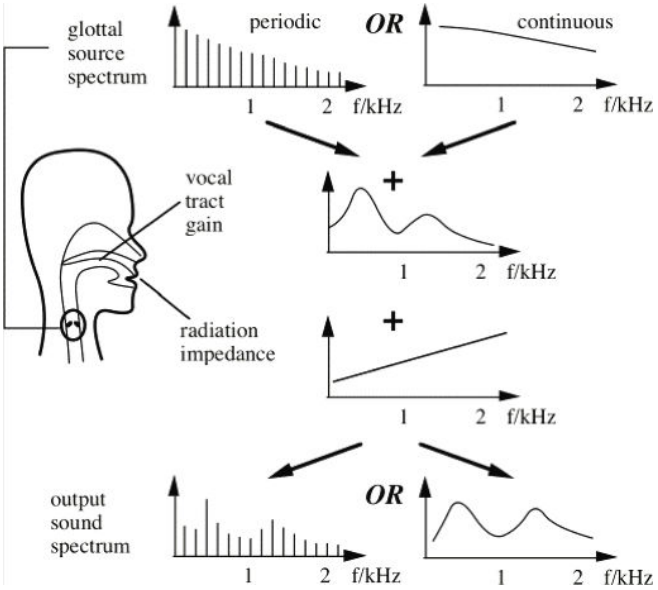


Figure 2-12 Source filter model for phonated speech. Reproduced with permission from Wolfe *et al.* (2009a). The glottal source spectrum (top) can be either periodic (harmonic spectrum) or broadband (continuous spectrum). The source is filtered by the gain of the vocal tract and the radiation impedance of the open mouth (centre), to produce the output sound spectrum (bottom). The y-axes show sound pressure level and so the + denotes that the impedances are multiplied. The formants (peaks in the output sound spectrum) occur at frequencies close to those of the vocal tract resonances, or peaks in the vocal tract gain function.

2.5.2.Limitations to the independent source-filter model

A theory of voice production in which the source and filter are independent cannot adequately explain subtleties of voice quality and extreme cases of what is now termed source-filter interaction. These interactions are categorised below.

2.5.2.1. Source influences on the filter

Source-filter interaction includes effects of the source on the filter. The properties of the vibrating vocal folds (e.g. glottal aperture and open quotient) influence the acoustic resonances of the vocal tract (Barney *et al.*, 2007; Lulich *et al.*, 2009; Swerdlin *et al.*, 2010).

2.5.2.2. Filter influences on the source

The models of the source discussed in 2.2.2 include some dependence between source and filter, as the acoustic load of the vocal tract is important. For example the simple mass spring model requires the presence of a vocal tract to oscillate (Flanagan, 1968). An inertive load upstream means that the airflow continues through the glottis, giving rise to the Bernoulli force. This is the behaviour highlighted by the Fletcher model (Fletcher, 1993; Tarnopolsky *et al.*, 2000).

The two-mass model, which is more frequently used today, demonstrates bifurcations and pitch jumps when f_0 and f_{R1} are comparable (Ishizaka and Flanagan, 1972; Pelorson *et al.*, 1994; Ruty, 2007) and is predicted to benefit from inertance downstream (Titze, 2008).

There is no strong experimental evidence for the impedance 'preferences' of the vocal folds, and they likely depend on the mode of oscillation. However, when f_{Ri} and nf_0 are close the vocal tract applies a large load impedance on the vocal folds, and the sign of the reactance changes across the resonance. This may be the cause of unintended effects reported in this range, such as pitch jumps or reduction in sound intensity (Hatzikirou *et al.*, 2006; Titze *et al.*, 2008). Additionally, the constriction of the aryepiglottic larynx affects f_0 (Bailly *et al.*, 2008; Lucero *et al.*, 2012; Ingo R. Titze, 2004; I. R. Titze, 2004; Titze, 2008; Titze and Story, 1997).

Note that the resonance tuning mentioned in 2.3.2, could in principle be explained by the independent source-filter model. However, since f_{R1} approaches f_0, some non-linear interaction may be present.

2.6. Straw Phonation: Possible Source-Filter Interaction

In speech therapy and voice training, various methods of occlusions or partial occlusions of the vocal tract are made, from placing the hand over the mouth, to lip trills, tongue trills, and their combination as 'blowing raspberries' (for an overview of the various methods, see Titze (2006)). Using straws to provide the partial occlusion allows for the aerodynamic and acoustic load on the vocal tract to be more precisely controlled than other methods by selecting the appropriate straw dimensions (Pillot-Loiseau *et al.*, 2009; Titze, 2002). Protocol for using the straws varies, but the effect is to raise the pressure in the vocal tract (the intra-oral pressure P_{io}), increase source filter coupling and to improve the vocal intensity and economy, i.e. using the minimum vocal fold collision necessary to produce the voice (Pillot-Loiseau *et al.*, 2009; Titze, 2006). Several physiological aspects of this technique have been studied on human subjects, such as the impact on laryngeal muscle activity and glottal adduction (Laukkanen *et al.*, 2008), and the articulatory and acoustical adjustments such as closing the velum (Laukkanen *et al.*, 2012; Vampola *et al.*, 2011). Further investigations have suggested that phonation into a straw could be a useful diagnostic tool to provide an estimation of phonation threshold pressure (Titze, 2009), or to provide an alternative to a voice range profile that avoids some of the difficulties associated with obtaining them (Titze and Hunter, 2011).

The physics of phonation into a straw is complex, since it changes both the DC and AC loads on the vocal folds. Further, human subjects can rapidly and unconsciously make adjustments to adapt their vocal gesture to changes in the phonatory situation (Baer, 1979). However, the implication of straw phonation as a therapeutic technique is that the person using the straw can learn to retain the changes after removing the straw. To simplify the problem, *in vitro* experiments provide a system to assess the underlying physics.

Using an *in vitro* model Bailly *et al.* (2008) showed that a vocal-tract constriction in the region of the aryepiglottic folds can either facilitate or impede vocal fold vibration by changing P_{io}. Chapter 5 describes experiments performed on the same model system, using latex vocal folds and a rigid vocal tract replica, with straws at the 'lips'. These focus on three aspects of the physics of straw phonation: the aerodynamic effects of the downstream constriction; the effect of the acoustic load provided by the straw, and the mechanical resonances of the vocal folds. These *in vitro* experiments suggest several ways that the straws may influence the vocal fold behaviour via source-filter interaction.

3 *In Vivo*: Energy Losses & Mechanical Resonances of the Vocal Tract

This chapter describes the *in vivo* application of the three-microphone three-calibration technique to the vocal tract to measure its acoustic and mechanical resonances with the glottis closed[5]. The measurements are used with a simple model of the vocal tract to determine the effective visco-thermal losses and the mechanical properties of the yielding walls.

3.1. Materials and Methods

3.1.1. The subjects

Ten volunteer subjects (seven men and three women) took part in the experiment. Approval for the use of human subjects was given by the ethics committee of the University of New South Wales, where all experiments were conducted. One subject (S6) was a professional singer employed by a national opera company, another (S9) was an amateur soprano. The remaining subjects were untrained in speaking or singing.

3.1.2. Measuring acoustic resonances of the vocal tract

Acoustic impedance measurements were made following the three-microphone three-calibration method of Dickens *et al.* (2007). A cylindrical aluminum measurement head was constructed with internal diameter 26.2 mm and microphones spaced at 20, 60, and 260 mm from the downstream measurement end. A broadband signal synthesised from sine waves between 10 Hz and 4.2 kHz (Smith, 1995) was injected via an enclosed 2" dome midrange loudspeaker (Jaycar CM-2092, Australia) at the upstream end. The signals from the three microphones (B&K 4944A, Brüel & Kjær, Nærum Denmark) were connected via a conditioning preamplifier (Nexus 2690, Brüel & Kjær, Nærum, Denmark) to a Firewire audio interface (MOTU 828, Cambridge, MA) and sampled at 44.1 kHz with 16 bit

[5] This work was performed under the supervision of John Smith and Joe Wolfe at UNSW. A study based on data from a single subject was published as 'Low frequency response of the vocal tract: acoustic and mechanical resonances and their losses,' Proceedings of the Australian Acoustical Society Conference, Fremantle, Australia, November 2012 (Hanna *et al.*, 2012a). Further preliminary portions of this work were presented in 'Resonances and bandwidths in the vocal tract and why they are important for speech comprehension,' Australian Institute of Physics Congress, Sydney, Australia, (Hanna *et al.*, 2012c); 'Acoustic measurements of vocal tract resonances,' International Conference on Voice Physiology and Biomechanics (ICVPB), Erlangen, Germany, (Hanna *et al.*, 2012b); 'Measurements of the aero-acoustic properties of the vocal folds and vocal tract during phonation into controlled acoustic loads,' International Congress on Acoustics (ICA), Montreal, Canada (Hanna *et al.*, 2013c); and 'The player-wind instrument interaction', Stockholm Music Acoustics Conference, Stockholm, Sweden (Wolfe *et al.*, 2013).

resolution. The upper frequency limit of the measurements is determined by the smallest spacing between any two of the three microphones, which here causes a singularity at 4.3 kHz. The lower frequency limit of around 10 Hz was determined by the frequency response of the loudspeaker and the microphones.

The impedance head was calibrated using three non-resonant acoustic loads: a quasi-infinite impedance (a large rigid mass), a purely resistive impedance (an acoustically infinite pipe), and a large flange (Dickens *et al.*, 2007). Measurements from the microphone array allow determination of the acoustic impedance Z at a reference plane, located in these experiments at the end of the cylindrical impedance head just inside each subject's lips.

With this measurement technique, the geometry on the loudspeaker side of the microphones does not affect the measurement. Here, for the measurements on subjects, a narrow tube (5 cm long and 7.8 mm diameter) filled with acoustic fibre was connected in parallel with the loudspeaker, which permitted the DC airflow necessary for phonation. A schematic of the apparatus is shown in Figure 3-1.

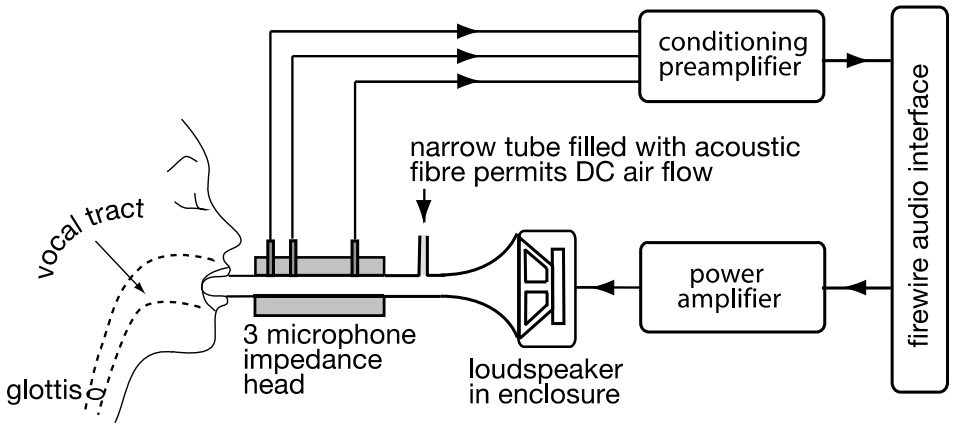


Figure 3-1 Schematic diagram (not to scale) showing how the vocal tract impedance is measured during phonation.

3.1.3.Experimental protocol

The subjects were asked to find a comfortable position for the impedance head in their mouth and to ensure an airtight seal with their lips around the outer diameter of 31.8 mm. For phonation, the subjects were asked to sustain the vowel in the word 'heard', which corresponds approximately to the vowel [ɜː] in the International Phonetic Alphabet, for the duration of the broadband signal. Subjects were given time to practice this unusual configuration, and some subjects chose to pinch their nose to avoid nasalization. At least

three measurements were made for each subject during each of the two gestures: one during a comfortable low pitched modal (mechanism 1) phonation comparable with their speaking voice, with the velum closed, and the other one while miming the vowel with the glottis and velum closed.

Since the carrier word ends in "rd", movement of the tongue would be expected to occur during its pronunciation. This effect was controlled in two ways: firstly the subjects were asked to focus on producing the vowel sound and not the entire word. Secondly, the initial and final cycles of the injected broadband signal were discarded prior to analysis to remove any transient parts of the measurement, including those due to the loudspeaker.

During each gesture, 8 or 16 cycles of the broadband signal (plus an initial and final cycle that were discarded) were injected through the measurement head into the subject's mouth. The measured impedance for each cycle was displayed in quasi real-time so that changes in the geometry of the vocal tract during measurements, particularly a clear change in the impedance spectrum corresponding to the opening of the velum, could be detected and the measurement repeated if necessary. Later, the approximate location of the maxima and minima of the impedance spectra were identified with a 6^{th} order Savitzky–Golay smoothing filter (Savitzky and Golay, 1964). Parabolas were fitted to both sides of each of the extrema and their crossover point used to determine the frequency f and the impedance magnitude $|Z|$. B was determined as the difference between the frequencies at half maximum power.

3.1.4.Adjustments for low frequency and during phonation

Measurements with the glottis closed were initially made using a single broadband spectrum from 14 to 4200 Hz. However, the low frequency components in the initial broadband signal excite mechanical resonances in the tissues surrounding the vocal tract. These resonances sometimes disturbed the subjects and caused them to alter their tract geometry during a measurement, or to produce an uncontrolled vibrato; in both cases this interfered with the measured impedance spectrum. For this reason the frequency range of interest was divided into three different broadband frequency ranges.

To determine the acoustic resonances, both with glottis closed and during phonation, a broadband signal of 200-4200 or 300-4200 Hz with a resolution of 2.69 Hz (44.1 kHz/2^{14}) was injected; this avoided the aforementioned unwanted effects of the injected low frequency components.

When studying the mechanical resonances, two narrow, low frequency windows were used: 10-50 Hz with a frequency resolution of 0.34 Hz (44.1 kHz/2^{17}), and 14-300 or 14-400 Hz with a resolution of 0.67 Hz (44.1 kHz/2^{16}). During these measurements four of the subjects also had a small magnet attached to their cheek and/or neck with adhesive to monitor the movement by inducing a voltage in a coil of wire placed on the axis a fixed distance away.

3.2. Results and Discussion

3.2.1. Acoustic resonances of the vocal tract

3.2.1.1. Impedance spectrum

Figure 3-2 and Figure 3-3 show typical acoustic impedance spectra of men and women with their glottis closed. The minima above about 500 Hz are the acoustic resonances *R*i that produce the formants *F*i. The maxima are anti-resonances *A*i, which in speech would correspond to minima in the spectral envelope of the radiated sound. Here they are given a corresponding index i. Ellipses in the figures represent the means and standard deviations of frequency *f*, bandwidth *B*, Q factor *Q* and impedance magnitude $|Z|$ for each resonance and anti-resonance for all subjects of the same sex.

In speech, f_{R1} typically has a value between about 300 and 800 Hz, the value increasing with mouth opening, while f_{R2} varies between about 800 and 2000 Hz, depending largely on tongue shape (Peterson and Barney, 1952). For the neutral vowel [əː] pronounced by a typical male subject $f_{R1} \sim f_{F1} \sim 500$ Hz, $f_{R2} \sim f_{F2} \sim 1500$ Hz, etc. The mean *f*, *B* and $|Z|$ of the resonances determined over the entire frequency range are given in Table 3-1. Note that in these experiments, the impedance head constrains the mouth aperture for all subjects, limiting the possible range of f_{R1}. Although subjects were asked to keep their tongue low, greater variation was observed in the frequencies of the higher acoustic resonances due to tongue movement and other unintentional changes to the vocal tract geometry.

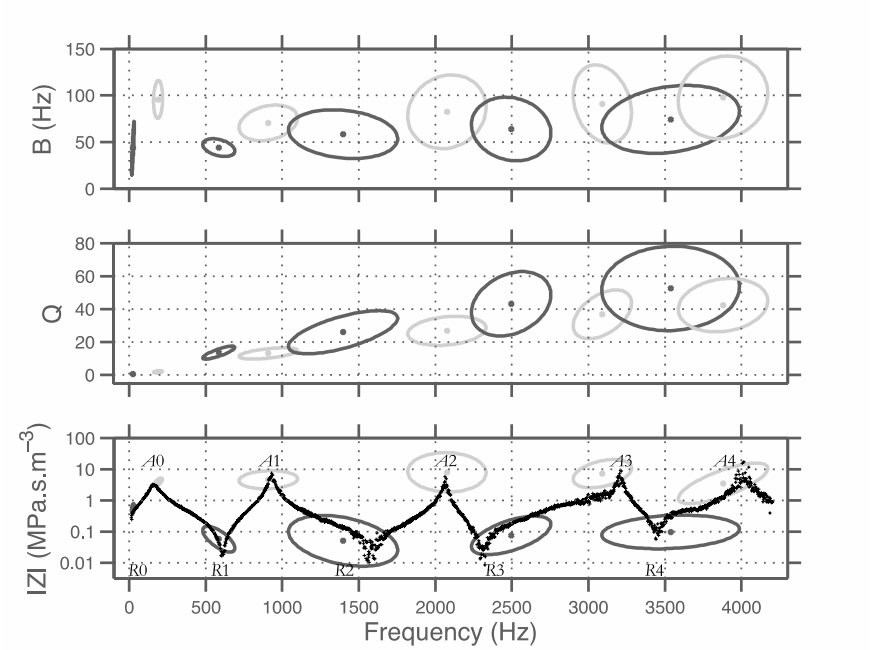


Figure 3-2 Dependence on frequency of the measured bandwidth, Q factor and impedance amplitude for each resonance *R*i (dark) and anti-resonance *A*i (pale) for male subjects with glottis closed. An example impedance magnitude spectrum is included in the lowest plot. The ellipse semi-axes indicate one standard deviation, with the major axis in the direction of correlation.

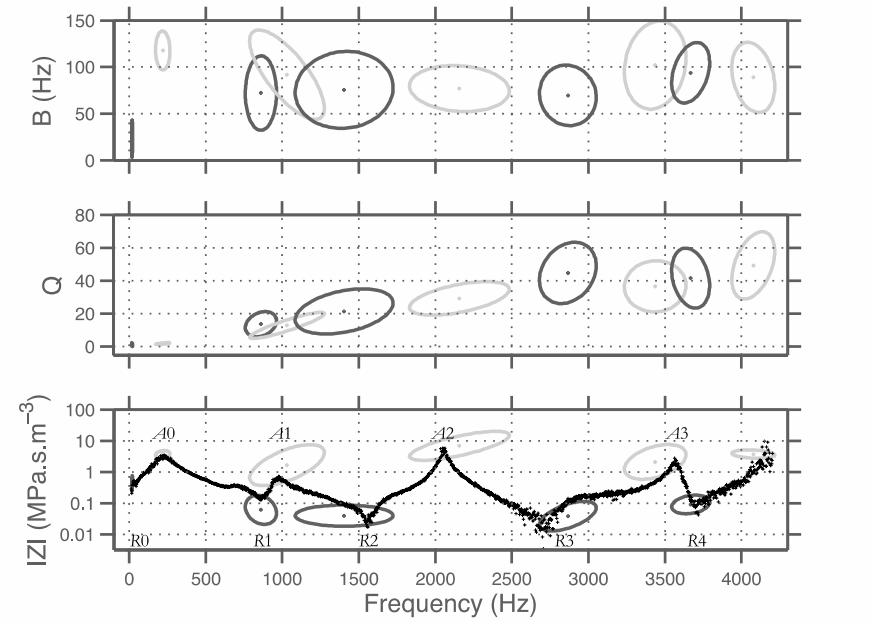


Figure 3-3 Dependence on frequency of the measured bandwidth, Q factor and impedance amplitude for each *R*i (dark) and *A*i (pale) for female subjects with glottis closed. Details as in Figure 3-2.

Table 3-1 Mean male (left) and female (right) data with closed glottis, and measurements during phonation (shown in italics). Phonation values are discussed in Chapter 4. Those marked with an asterisk are significantly different to the closed glottis values at the $p < 0.01$ level. Note that *A*4 is not included since the impedance maximum is not always within the measured frequency range, 10-4200 Hz. No. Samples is the number of cycles that displayed a clear minimum or maximum in the impedance spectrum associated with the resonance or anti-resonance. Not all extrema were clearly defined in each cycle of each measurement.

Male	*f* (Hz)		*B* (Hz)		|Z| (MPa·s·m^{-3})		No. samples	
*R*0	25±5	*25±10*	40±20	*20±10**	0.55±0.15	*0.39±0.12**	184	*100*
*A*0	210±55	*245±35**	105±30	*85±30**	3.9±1.0	*2.6±0.96**	250	*188*
*R*1	585±70	*570±55*	45±5	*85±30**	0.069±0.051	*0.15±0.11**	221	*113*
*A*1	940±155	*1010±175**	70±10	*125±30**	8.3±10	*4.2±4.6**	229	*146*
*R*2	1415±250	*1390±200*	60±15	*90±25**	0.12±0.15	*0.17±0.17**	226	*119*
*A*2	2075±170	*2015±175**	80±25	*100±35**	12±11	*6.2±3.5**	218	*142*
*R*3	2465±190	*2425±120*	65±20	*85±20**	0.13±0.12	*0.22±0.23**	221	*138*
*A*3	3105±285	*3005±215**	100±35	*105±35*	11±0.65	*11±7.9*	214	*138*
*R*4	3535±295	*3620±370*	75±25	*95±20**	0.13±0.12	*0.18±0.13**	192	*90*
Female	*f* (Hz)		*B* (Hz)		|Z| (MPa·s·m^{-3})		No. samples	
*R*0	20±5	*20±5*	30±15	*15±5**	0.48±0.18	*0.43±0.11*	47	*22*
*A*0	215±30	*245±25**	115±15	*105±35**	3.7±0.7	*2.7±0.8**	121	*52*
*R*1	860±65	*800±65**	70±25	*75±30*	0.077±0.057	*0.28±0.16**	90	*43*
*A*1	1035±160	*970±165**	90±30	*110±45**	2.4±1.7	*1.5±0.6**	97	*73*
*R*2	1405±210	*1515±95**	75±25	*90±25**	0.045±0.024	*0.071±0.063**	118	*71*
*A*2	2160±215	*2340±125**	75±15	*115±35**	9.0±7.3	*4.0±3.4**	121	*77*
*R*3	2895±150	*2625±180**	75±25	*90±20**	0.077±0.12	*0.18±0.06**	97	*80*
*A*3	3445±135	*3495±145*	100±30	*100±25*	0.29±0.25	*9.3±5.8**	107	*71*
*R*4	3710±150	*3955±200**	95±20	*90±20*	3.8±0.71	*0.31±0.22**	76	*51*

3.2.1.2. Comparing the vocal tract with a rigid cylinder

A simple cylindrical model of the tract shown in Figure 3-4 provides a useful comparison with the measurements presented in Figure 3-2, Figure 3-3 and Table 3-1. For frequencies below a few kilohertz, wavelengths are much greater than the transverse dimensions of the vocal tract, so a simple one-dimensional approximation of the vocal tract as a rigid tube captures much of the physics. Curvature of such a tube has little effect in the frequency range studied (Kob, 2002; Sondhi, 1986), so is neglected in this analogy. At higher

frequencies, deviations from cylindrical geometry become more important, and at frequencies below a few hundred hertz, the non-rigidity of the walls needs to be considered (Makarov and Sorokin, 2004; Sondhi, 1974) as discussed in section 3.2.2.

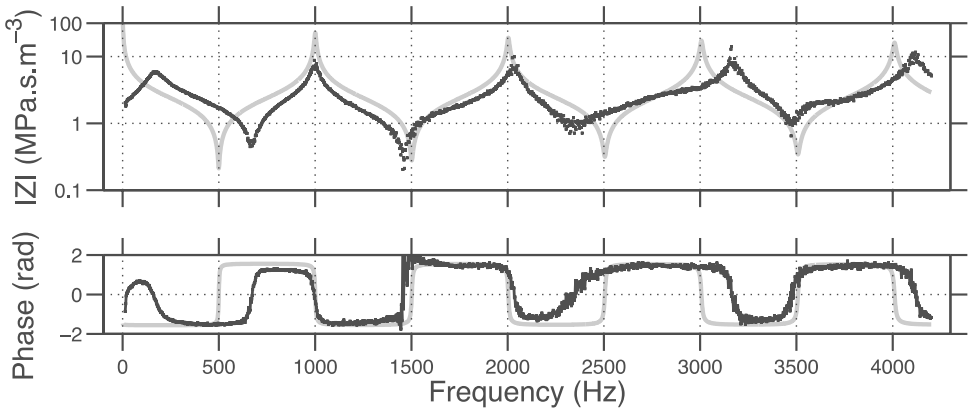


Figure 3-4 Typical impedance spectrum (magnitude and phase) for one male subject measured through the lips, with the glottis closed (dark), compared with the calculated impedance of a closed rigid pipe of similar dimensions (examples for each subject are given in Appendix A). Data were measured over a single cycle of the broadband signal lasting 370 ms. The low frequency data from 10 to 400 Hz are shown in more detail in Figure 3-6.

The impedance Z measured through the lips is analogous to the impedance of a rigid cylinder loaded by an impedance Z_L as described in 2.3.1. Z is given by the equation (2-2). The wavenumber Γ is a complex number as defined in equation (2-3), which includes an attenuation coefficient α that accounts for losses to the tube walls by viscous drag and thermal conduction. The effect of α on the impedance spectrum is discussed in section 3.2.1.3.

For a vocal tract with closed glottis, the impedance measured at the lips is similar to Z_{closed} in equation (2-4), which has minima at the resonances Ri. These occur at $\lambda_{Ri} = 4l, 4l/3, 4l/5$ etc. much like the evenly spaced formants of the neutral vowel [ɘː], which occur in the ratio 1:3:7 etc. Impedance maxima, Ai, occur at $\lambda_{Ai} = 2l, 2l/2, 2l/3$. An effective vocal tract length l can therefore be estimated from the measured f_{Ri} and f_{Ai} in Table 3-1 as the arithmetic mean of the length determined by these relationships using the speed of sound c:

$$l = \frac{1}{7}\left(\frac{c}{4f_{R1}} + \frac{c}{2f_{A1}} + \frac{3c}{4f_{R2}} + \frac{c}{f_{A2}} + \frac{5c}{4f_{R3}} + \frac{3c}{2f_{A3}} + \frac{7c}{4f_{R4}}\right) \quad (3\text{-}1)$$

This yields average lengths of 173 mm and 155 mm for men and women, respectively. The lengths are slightly longer but the differences between sexes are comparable to the reported 1.3 cm difference between the lengths of the tracts of post-pubertal men and women obtained from a study of MRI data (Fitch and Giedd, 1999).

Three of the male and two of the female subjects were comfortably able to put their teeth and lips around the impedance head, the remaining subjects sealed it with their lips only. No systematic difference was observed in the overall form of the impedance spectra and the subjects' kept their chosen position throughout the experiment, with the standard deviation of each subject's f_{R1} typically less than 10%.

The acoustic impedance spectrum (magnitude and phase) shown in Figure 3-4 is for a rigid walled cylinder that is closed at the far (glottis) end, with length l = 170 mm and radius r = 10 mm (values from the average male data determined in section 3.2.2) calculated using equation (2-4). A qualitative difference between this simple model and the vocal tract is that, for frequencies below the first acoustic resonance *R*1, the impedance of a closed rigid tube increases with decreasing frequency – monotonically to infinity. The vocal tract is not rigid, so the resonance and anti-resonance below f_{R1} in the measured impedance spectra are attributed to mechanical properties of its walls, which are discussed in section 3.2.2.

3.2.1.3. Bandwidth and quality factor

Figure 3-2, Figure 3-3 and Table 3-1 display the mean bandwidth *B* calculated for the experimental subjects. For measurements with the glottis closed an unpaired t-test shows that *B* for each of the extrema differ at the $p < 0.01$ level. The bandwidths of the acoustic resonances and anti-resonances (those above 300 Hz) increase with increasing frequency at a rate of ~1 Hz/100 Hz (*B* increases from ~50-90 Hz over the frequency range measured) for men and half that (*B* ~70-90 Hz) for women, over the frequency range measured.

The bandwidths of a rigid tube with comparable dimensions are significantly narrower and the Q factors higher than those of the closed vocal tracts (note the sharper extrema of the model curve in Figure 3-4). For a rigid cylinder, the losses that determine *B* are dominated by visco-thermal energy losses at the boundary layer, represented by the attenuation coefficient α, number, as shown in equation (2-3). For the vocal tract, the attenuation coefficient is expected to be higher than in a rigid, dry cylinder of comparable dimensions for at least the following reasons: *i)* it has a larger surface area:volume ratio, *ii)* it is not completely rigid and *iii)* it has a wet surface. (Humidity has been shown to increase losses

(Coltman, 2003)). In principle, each of these has a different frequency dependence, and each involves unknown quantities. However, the net effect of such energy losses is to decrease $|Z|$ and to increase B at Ri and similarly to increase $|Z|$ and to decrease B at Ai. So to improve the agreement between the model and the vocal tract, Figure 3-5 shows the effect on the impedance and associated bandwidths of the model when the attenuation coefficient α is increased by a factor of five to produce a curve whose extrema have magnitudes and bandwidths comparable with those of the male *in vivo* data in Table 3-1. Note that, unlike in Figure 3-4, in Figure 3-5 the tract is modelled with non-rigid parameters as described in section 3.2.2. This has only a minor effect on the impedance curve in the range of the acoustic resonances.

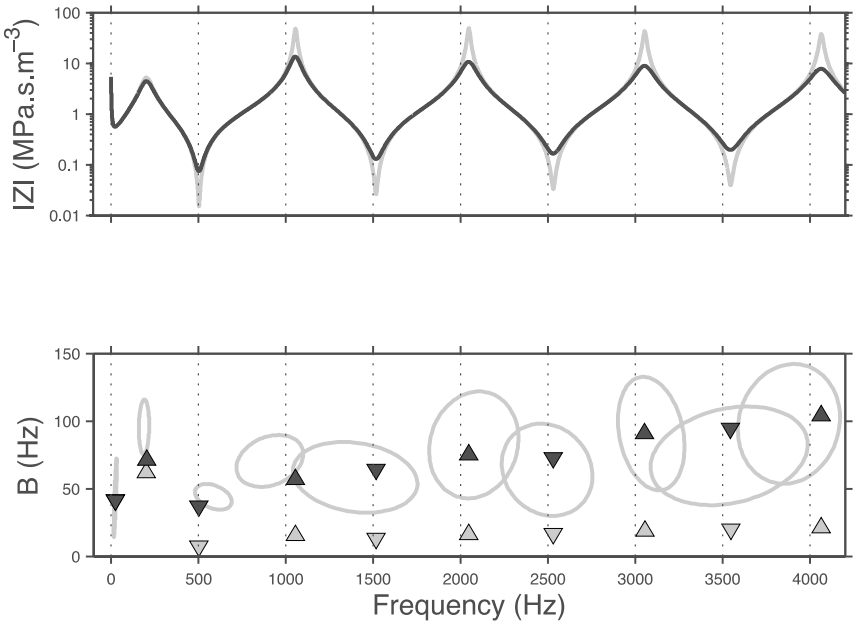


Figure 3-5 (top) Calculated impedance magnitude for a non-rigid vocal tract model with parameters from Table 3-2 (pale) *cf.* rigid model in Figure 3-2, discussed below. The dark line shows the same model with visco-thermal attenuation coefficient α increased by a factor of five. (bottom) Triangles show B_{Ri} and B_{Ai} for the two models. Filled triangles show α increased by a factor of five to be comparable with the measured male B ellipses (reproduced from Figure 3-3). The infinitely rigid model does not have the mechanical resonance and anti-resonance $R0$ and $A0$.

3.2.2.Mechanical resonances and the low frequency model

At frequencies below f_{R1} two additional low frequency extrema, $R0$ and $A0$, are present in the impedance spectra measured at the lips (Figure 3-2, Figure 3-3, Figure 3-4 and Figure 3-5). These do not appear in the spectra of a rigid tube, but can be included in the simple model by considering the finite mass and rigidity of the walls, with a few assumptions. The maximum in impedance $|Z|_{A0}$ corresponds to the oscillation of the walls on the 'spring' of the air in the tract. The minimum in impedance $|Z|_{R0}$ is the result of the tissue mass oscillating on its own stiffness. Again, the vocal tract is compared with a uniform cylinder with walls of surface area S, and uniformly distributed mass m and thickness w. At frequencies well below f_{R1}, the dimensions of the tract are much less than the wavelength, so that the tract may be approximated as an acoustically compact object, i.e. one in which the pressure acts uniformly. So, for these frequencies, pressure acting on the surface area S of the tract walls accelerates tissue with an acoustic inertance

$$(3\text{-}2) \qquad L_t = m/S^2$$

in series with an acoustic compliance

$$(3\text{-}3) \qquad C_t = S^2/k$$

due to the deformability of the walls. These two elements are in series because the pressure-induced vibration of the walls involves the same acoustic current in both. The biological tissue of these walls is further approximated as having a spring constant $k = -2\pi r L_t \mathrm{d}P/\mathrm{d}r$ for hypothetically uniform radial dilation $\mathrm{d}r$ due to a homogenous pressure change $\mathrm{d}P$. Additionally, since an increase in internal pressure also compresses the air within the tract, the compliance of the air C acts in parallel with the series circuit just mentioned. An equivalent circuit for this mechanical model of the tract at very low frequencies is shown in Figure 3-6, which also shows the measured impedance spectrum for the same subject as Figure 3-2 in the low frequency range. For comparison, an example spectrum of the same subject inhaling and phonating are included. These are discussed in chapter 4. As previously noted, the measurements made with phonation in this range were affected by the broadband signal and so are not clear.

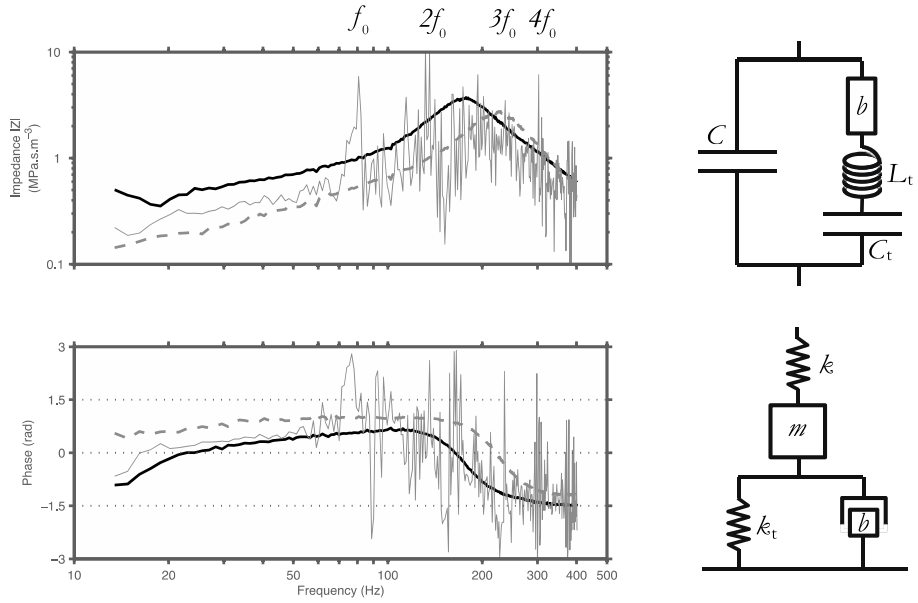


Figure 3-6 (Left) Impedance spectra plotted on a log scale showing low frequency mechanical resonances for closed glottis condition (dark), inhalation (pale dashed) and phonation (pale) (examples for each subject are given in Appendix A). The first few harmonics of the phonation spectrum are indicated as f_0, $2f_0$ etc. The inhalation and phonation measurements are discussed in chapter 4. The low frequency mechanical resonance was close to the low frequency limit of the measurements but was evident in the measurements on all subjects for the closed glottis and phonation conditions. (Right) Equivalent electrical (top) and mechanical (bottom) analogues.

To account for the losses accompanying the tissue motion, a resistor b is positioned in series with L_t and a dashpot b in parallel with k and $(m+k_t)$. Use of a resistor or dashpot retains the linearity of the model. However, it is unlikely that the losses in the tissue are accurately modelled by a purely viscous term.

Including these mechanical elements and parameters into the simple vocal tract model introduces the desired additional low frequency series mechanical resonance $R0$, and parallel anti-resonance $A0$, in addition to the acoustic resonances $R1,2,3,\ldots$ and anti-resonances $A1,2,3,\ldots$ Similar electrical analogues have previously been proposed to describe the expected mechanical resonances of the tract (a summary of this approach is given by Wakita and Fant (1978)). However they have typically lacked some of the parameters necessary to define the system completely. Makarov and Sorokin (2004) determined that using uniform distributed values for the wall parameters is no less accurate for the purpose of determining formant frequencies than separating the tract into regions of varying wall mass, thickness etc.

Since $f_{A0} \sim 10 f_{R0}$, the anti-resonance $A0$ and resonance $R0$ can be treated separately for the purpose of estimating parameters: at low frequency (i.e. evaluating $R0$), C is approximately an open circuit, while at high frequency (i.e. evaluating $A0$) C_t is a short circuit. Because the same tissue is involved in both cases, the same resistance b and wall inertance L_t are used for the parallel and series resonances, respectively. The mechanical properties can then be estimated from the measured values of f, Q and $|Z|$.

For the series resonance f_{R0} depends on L_t and C_t.

$$f_{R0} = \frac{1}{2\pi\sqrt{L_t C_t}} \quad (3\text{-}4)$$

For the parallel resonance, f_{A0} is measured at the maximum impedance and is slightly lower than the characteristic frequency f of the circuit due to the finite Q. In practice, there is little difference between these frequencies; nonetheless the two frequencies are related in the following way (Langford-Smith, 1963, pp. 150–151) :

$$f = \frac{1}{2\pi\sqrt{L_t C}} = f_{A0} \Big/ \sqrt{\frac{-1}{{Q_{A0}}^2} + \sqrt{\frac{2}{{Q_{A0}}^2} + 1}} \quad (3\text{-}5)$$

This characteristic frequency and the measured $|Z|_{A0}$ determine C, since

$$C = Q_{A0}/2\pi f |Z|_{A0} \quad (3\text{-}6)$$

and hence L_t from equation (3-5).

The acoustic compliance C is due to a compact volume V

$$V = C\gamma P_A \quad (3\text{-}7)$$

where γ is the adiabatic constant and P_A the atmospheric pressure. These acoustic components are related to tissue parameters by geometry. For the approximately cylindrical vocal tract, the effective radius

$$r = \sqrt{V/\pi l} \quad (3\text{-}8)$$

where l is the effective length of the tract from equation (3-1). So, the surface area of the tract walls

$$S = 2\pi rl \quad (3\text{-}9)$$

which can be used to determine the effective mass of the walls

$$m = S^2 L_t \quad (3\text{-}10)$$

and their thickness w, using a tissue density $\rho \sim 10^3$ kg.m^{-3}.

$$w = SL_t/\rho \quad (3\text{-}11)$$

Returning to the series resonance $R0$, the compliance of the tissue C_t is calculated from equation (3-4) using the resonance frequency f_{R0}, which determines the tissue spring constant

$$k = S^2/C_t \quad (3\text{-}12)$$

Finally, the resistance terms in both circuits are given by the relations

$$b_{A0} = 2\pi f L_t/Q_{A0} = \frac{\sqrt{L_t/C}}{Q_{A0}} \quad (3\text{-}13)$$

$$b_{R0} = 2\pi f_{R0} L_t/Q_{R0} = \frac{\sqrt{L_t/C_t}}{Q_{R0}} \quad (3\text{-}14)$$

The values determined from these calculations are given in the first two columns of Table 3-2. The derivation of other terms follows below.

Table 3-2 First two columns: mechanical resonance parameters measured (upper in bold) and derived (bottom) in this study. For comparison, subsequent columns show data reported by Fant *et al.* (1976) and Ishizaka *et al.* (1975). Italics show values that include assumptions detailed in the text. The numbers of male and female subjects are given in parentheses.

			Fant *et al.*		Ishizaka *et al.*	Male (1)	
	Male (7)	Female (3)	Male (5)	Female (7)	Relaxed cheek	Tense cheek	Neck
f_{R0} (Hz)	**25**	**19**	*25*	*19*	**32**	**60**	**72**
f_{A0} (Hz)	**212**	**216**	**191**	**218**	184	218	170
$\|Z\|_{R0}$ (MPa·s·m^{-3})	**0.55**	**0.48**	0.22	0.18	**0.72**	**0.95**	**2.1**
$\|Z\|_{A0}$ (MPa·s·m^{-3})	**3.9**	**3.7**	**3.0**	**4.5**	4.5	3.8	4.9
B_{R0} (Hz)	**41**	**28**	32	20	**62**	**115**	**153**
B_{A0} (Hz)	**104**	**116**	**76**	**94**	86	103	80
Q_{R0}	**0.80**	**1.0**	*0.8*	*1.0*	**0.52**	**0.52**	**0.47**
Q_{A0}	**2.1**	**1.9**	**2.5**	**2.3**	*2.1*	*2.1*	*2.1*
L_t (kg·m^{-4})	1.4 x10^{3}	1.4 x10^{3}	**1.1 x10^{3}**	**1.4 x10^{3}**	1.9 x10^{3}	1.3 x10^{3}	2.1 x10^{3}
C (m^3/Pa)	4.1 x10^{-10}	3.7 x10^{-10}	**6.9 x10^{-10}**	**3.7 x10^{-10}**	*4.1 x10^{-10}*	*4.1 x10^{-10}*	*4.1 x10^{-10}*
V (m^3)	5.8 x10^{-5}	5.2 x10^{-5}	**9.8 x10^{-5}**	**5.3 x10^{-5}**	*5.8* x10^{-5}	*5.8* x10^{-5}	*5.8* x10^{-5}
l (m)	0.17	0.16	*0.17*	*0.16*	*0.17*	*0.17*	*0.17*
r (m)	0.010	0.010	0.013	0.010	0.010	0.010	0.010
S (m^2)	0.011	0.010	0.015	0.011	*0.011*	*0.011*	*0.011*
m (kg)	0.17	0.15	0.23	0.16	**0.23**	**0.17**	**0.27**
w (m)	0.02	0.01	0.016	0.015	0.021	0.015	0.024
C_t (m^3/Pa)	2.9 x10^{-8}	4.9 x10^{-8}	3.6 x10^{-8}	5.0 x10^{-8}	1.3 x10^{-8}	3.4 x10^{-8}	2.3 x10^{-8}
K (kN·m^{-1})	4.3 x10^{3}	2.1 x10^{3}	5.9 x10^{3}	2.3 x10^{3}	**9.5 x10^{3}**	**3.7 x10^{3}**	**5.5 x10^{3}**
b_{R0} (MPa·s·m^{-3})	0.27 x10^{5}	0.18 x10^{5}	2.2 x10^{5}	1.8 x10^{5}	**7.2** x10^{5}	**9.5** x10^{5}	**21** x10^{5}
b_{A0} (MPa·s·m^{-3})	0.86 x10^{5}	1.0 x10^{5}	**5.3 x10^{5}**	**8.3 x10^{5}**	10 x10^{5}	8.4 x10^{5}	11 x10^{5}

The maximum in impedance $|Z|_{A0}$ corresponds to the parallel resonance of L_t and C. Q_{A0} is low, which is likely associated with the damping caused by tissue deformation rather than loss in the 'spring' of the air. The minimum in impedance of $|Z|_{R0}$ is the result of the series resonance of L_t and C_t. $Q_{R0} \sim 1$, suggesting that the spring of the tissue vibration is strongly damped. The equivalent circuit shows only one lossy component b. If the resistance acted as a viscous loss then $b_{R0} = b_{A0.}$ However, these values differ by a factor of ~3 for men and ~5 for women, which implies that the resistance term increases with frequency and/or amplitude of vibration. This frequency dependence is not explained by considering the losses as viscous. Nonetheless, the effective viscosity derived from these resistances, $\eta = w\rho/2\pi r^3$ = 90 and 500 Pa·m^{-2} (0.9 10^3 and 5.0 10^3 dyne/cm^2), respectively, are of the same order of magnitude as the scarce data in the literature of 75 – 200 Pa·m^{-2} for soft tissue in the subglottal tract (Habib *et al.*, 1994; Suki *et al.*, 1993).

The male L_t, C_t and mean b values from Table 3-2 are included in parallel with the modified cylindrical vocal tract model from the previous section to calculate the impedance spectrum in Figure 3-5. This model now includes the mechanical resonances, with values of f, $|Z|$ and B chosen to give values comparable to the measured impedance. Note that using one or the other value of b would give better agreement with the associated $|Z|$ and B.

The mechanical resonances were perceptible by the subject when excited by the low frequency probe signal. Measurement of the voltage induced in a coil of wire showed a distribution of vibration around the face and neck that was broadly in agreement with the contour plot produced by Fant *et al.* (1976) (reproduced in Figure 2-11) by supplying an acoustic probe signal through the lips to a closed vocal tract. Figure 3-7 shows the induced voltage in a coil of wire from a small magnet attached to the cheek during broadband excitation.

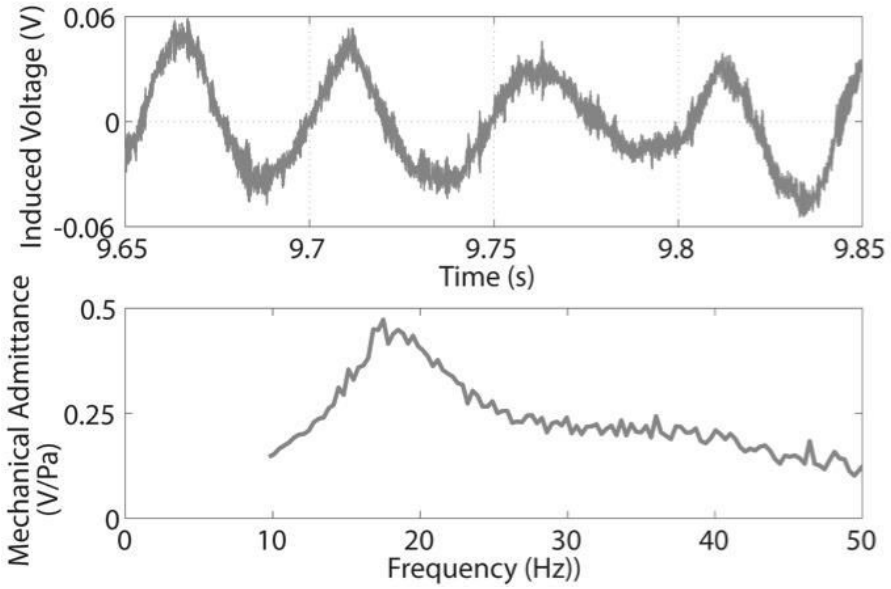


Figure 3-7 An electromotive force induced in a coil held 5 cm from a male subject's cheek shows the non-rigid behaviour of the vocal tract in the time domain, and as mechanical admittance in the frequency domain.

In the same study the mean maximum wall vibration frequency of five men and seven women was reported to occur close to the f_{A0} found in the current study. These data are shown in Table 3-2 along with the estimated L_t and V and hence one can infer C from equation (3-7). These results are also consistent with the lower limit of formant

measurements (Fujimura and Lindqvist, 1971), and estimates of the lowest resonance in speech under the elevated air pressure conditions experienced by divers (Fant and Sonesson, 1964).

It is possible to compare their data with the *A*0 data from the current study by making some assumptions about the tract parameters. Firstly, if the total inertance L_t corresponds to the total volume V of a single cylindrical cavity, the compliance of the air can be determined from equation (3-5). $|Z|_{A0}$ then follows from equation (3-6) and b_{A0} is given by equation (3-13). These data are in good agreement in the two studies, particularly for female subjects.

A further set of assumptions allows estimates of *R*0, which was not measured in the Fant *et al.* study. Assuming that the length of the tract is the same as the average length *l* from the current data, and that the tract is cylindrical, the effective surface area *S*, mass *m* and thickness *w* of the walls are determined from equations (3-9) and (3-11). Finally, if the resonance of the walls occurs at the frequency f_{R0} measured in the current study, with the current measured Q_{R0}, then the data can be extrapolated to determine the tissue compliance C_t, spring constant *k*, and resistance b_{R0} from equations (3-4), (3-12) and (3-14). And since the imaginary part of the series resonance *R*0 is zero at f_{R0}, then $|Z|_{R0} = b_{R0}$. Again, the values thus derived are in reasonable agreement with the data from the current study.

For *R*0, direct comparisons with the literature are less straightforward. The mechanical impedance of the cheek and neck of a single subject has been determined directly by external excitation with a shaker and used to calculate the equivalent *m*, *k*, and *b* per unit area (Ishizaka *et al.*, 1975). These data are also compared with the present study in Table 3-2 under the assumption that the vibration affects the same vocal tract surface area *S* used here, and $Z_{acoustic} = Z_{mechanical}/S^2$.

So, the mass *m* can be converted to an inertance L_t acting over the surface area of the tract *S* with equation (3-2). The compliance of the tissue is therefore given by equation (3-4), and the thickness can be determined by equation (3-11). The imaginary component of the impedance is zero at resonance, so the magnitude of mechanical impedance at resonance is equal to the resistance b_{R0} and therefore $|Z|_{R0} = b_{R0}/S^2$. A further assumption that the effective length and therefore volume of the tract are the same as those determined in the current study yields *C* and hence determines a hypothetical f_{A0},

which was not measured in that study, from equation (3-5). Finally, assuming that Q_{A0} is that measured in the current study, $|Z|_{A0}$ and b_{A0} follow from equations (3-6)and (3-13).

In all three excitation cases (relaxed cheek, tense cheek, and neck) f_{R0} is significantly higher and the bandwidths are significantly larger than those measured in the current study, perhaps due to differences between relatively localised excitation with a shaker and the uniform excitation by internal air-pressure. Despite these differences, the effective $|Z|$, m, k, and b values for the cheek are of the same order of magnitude as those presented here.

Finally the m, k and b determined in this study are also comparable with those quoted "according to various sources" by Makarov and Sorokin (2004) of 0.12-0.23 kg, 0.013-13 kNm^{-1} and 0.48-0.98 $MPa{\cdot}s{\cdot}m^{-3}$, respectively.

3.3. Summary

A new method yields direct measurements of the frequency, magnitude and bandwidths of vocal tract resonances over a range of nine octaves.

For an articulation approximating the neutral vowel [ɜː] and with glottis closed the impedance magnitudes of purely acoustic vocal tract resonances were 0.1 $MPa{\cdot}s{\cdot}m^{-3}$ and anti-resonances 10 $MPa{\cdot}s{\cdot}m^{-3}$. The bandwidths of the resonances up to 4 kHz in the absence of radiation loss from the mouth increase with frequency from approximately 50-90 Hz for men and 70-90 Hz for women. Comparison of the measured data with a simple one-dimensional rigid tube gives good agreement between 300 and 4200 Hz if the attenuation coefficient is five times higher than the typical visco-thermal wall losses of a rigid wall.

At lower frequencies, the finite mass and rigidity of the vocal tract walls strongly affect the acoustics of the tract. A low Q impedance minimum due to the mechanical resonance of the tissue walls of the vocal tract was observed at ~20 Hz. An impedance maximum at ~200 Hz is attributed to a resonance associated with tissue mass and the compliance of the air inside the tract. In this compact component, uniformly distributed model, the effective vibrating masses of the vocal tract walls are 0.24 and 0.14 kg for men and women respectively, with a spring constant of 4.3 and 2.3 kNm^{-1}.

4 *In Vivo*: The Glottis and Subglottal Vocal Tract

The *in vivo* vocal tract acoustic impedance measurement technique introduced in the previous chapter is used to make measurements with the glottis open: during phonation and inhalation. A simple acoustic model is used to calculate the glottal impedance, the impedance of the subglottal tract, and their influence on the impedance that drives the vocal folds[6].

4.1. Materials and Methods

The acoustic impedance measurements were made with the same subjects and during the same sessions as those in the previous chapter. The subjects were asked to sustain the vowel in the word 'heard' in a low pitched, modal voice for the duration of the broadband signal, or to inhale deeply while keeping the same vocal gesture for the duration of the signal.

4.2. Results and Discussion

The results are discussed with regard to modifications to the non-rigid vocal tract model presented in chapter 3. The load impedance on the vocal tract is the air in the glottis, which is assumed to operate as a compact object in series with the subglottal tract. So, when the glottal impedance is high, coupling to the subglottal tract is negligible. When the glottal impedance is low, the impedance of the subglottal tract must be considered.

4.2.1. High glottal impedance

4.2.1.1. Measurements during phonation

The rapid opening and closing of the glottis during phonation affects the measured impedance in several ways. Most conspicuously, the measured impedance spectra include superimposed signals at the fundamental phonation frequency f_0 and its harmonics $2f_0$, $3f_0$, etc. with a phase uncorrelated with the relevant components of the broadband signal as shown in Figure 4-1.

[6] This work was performed under the supervision of John Smith and Joe Wolfe at UNSW. Some examples of measurements with different glottal apertures were presented in 'Measurements of the aero-acoustic properties of the vocal folds and vocal tract during phonation into controlled acoustic loads,' International Congress on Acoustics (ICA), Montreal, Canada (Hanna *et al.*, 2013c) and 'The player-wind instrument interaction', Stockholm Music Acoustics Conference, Stockholm, Sweden (Wolfe *et al.*, 2013).

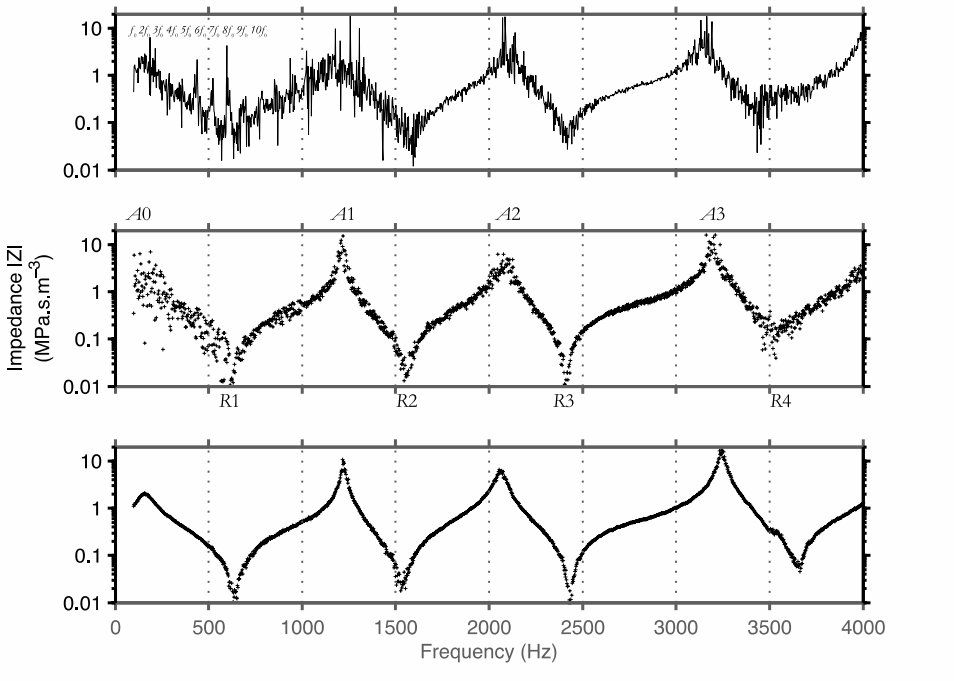


Figure 4-1 Three impedance spectra for one subject phonating with f_0 = 88 Hz in one gesture (top). Superimposed over the acoustic impedance (the broadband signal) are the harmonics of the voice, the first several of which are indicated. Unlabelled harmonics are visible up to ~2 kHz. $8f_0$ coincides with f_{R1} and this artefact obscures the impedance minimum. The middle plot shows a later cycle during which the phonation ceased, and the lower plot shows a cycle after phonation had ceased. The resonances *R*1-4 and anti-resonances *A*0-3 are indicated on the middle plot. Note that the sharpness of the peaks increases (i.e. *B* decreases) when phonation ceases as a consequence of the closed glottis and that the frequencies of some of the resonances change slightly. Examples of closed glottis and phonation measurements for all subjects are shown in Appendix A.

Disregarding the harmonics of the voice, the proportional changes to the impedance spectra are summarised in Figure 4-2. Their significance was assessed using an unpaired t-test and differences with $p < 0.01$ are indicated with filled bars in the figures (and with an asterisk in Table 3-1). For the grouped male data, changes in the frequencies of the acoustic *R*i are insignificant at the 1% level, which means that the closed glottis configuration preserves the resonance structure (and thus the vowel) of the phonation measurement. The frequency changes of the anti-resonances are significant, but they differ by small amounts. Only the change in f_{A0} is both significant and large (>10%). Bandwidth and magnitude changes for all resonances and anti-resonances, with the exception of *A*3, are significant at the 1% level. Considering the impedance of the glottis during phonation as an effective mass of air between the vocal folds, the inertive reactance of this mass would increase with

increasing frequency. This would explain why the acoustic resonances (at high frequency) are less affected by the change from closed glottis to phonation than the low-frequency acoustic-mechanical resonance.

Trends in the female data are less clear, particularly for frequency changes, but appear to follow the same patterns for B and $|Z|$.

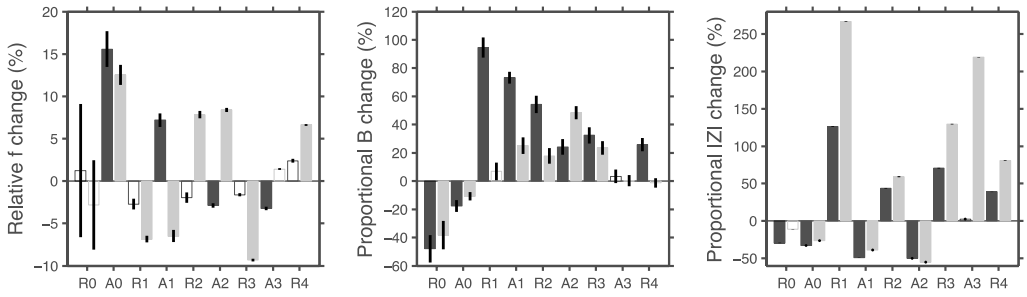


Figure 4-2 Proportional change in: frequency (left), bandwidth (middle) and impedance magnitude (right), from closed glottis to phonation, i.e. (*f*(Phonation)–*f*(closed))/*f*(closed). Filled bars show significant changes for males (dark) and females (pale) at the $p < 0.01$ level. The standard deviation is shown as a black vertical line.

Pham Thi Ngoc and Badin (1994) reported that changes in resonance frequencies from closed glottis to phonation were negligible, except in the case of the closed French vowels [i] and [y], which increase by more than 10% during phonation. They suggested that this effect could be due to differences in tongue position due to absence of feedback from the airflow when the glottis was closed. However, as discussed in 4.2.1.4.3 below, the lower limit of f_{R1}, which is approached in these closed vowels, must occur at a higher frequency than f_{A0}, which here increased by ~15% from closed glottis to phonation, and so it is then plausible that f_{R1} may increase in frequency by a similar amount for vowels with low f_{F1}.

4.2.1.2. Acoustic model of the vocal tract

The cylindrical model of the vocal tract with enhanced attenuation coefficient α, in parallel with the low frequency b, L_t, C_t circuit from the previous chapter, is consistent with the measurements made at the lips if the load impedance Z_L in equation (2-2) is replaced with a finite glottal impedance as described below.

4.2.1.2.1. The impedance of the glottis

During inhalation the vocal folds are abducted to provide a relatively large fixed glottal aperture. The geometry of the airway in the vicinity of the vocal folds is complex (Šidlof *et*

al., 2008) but to retain the simplicity of the model, the acoustic impedance of the glottis Z_g is treated as a cylinder of length l_g (including end effects) with a radius r_g that is smaller than that of the vocal tract. Its impedance is given by equation (2-2), where Z_L is the load of the impedance of the subglottal tract discussed below.

The rapid oscillation of the vocal folds during phonation presents a smaller but non-zero average opening. Hoppe *et al.* (2003) report a maximum glottal area during low pitched, modal phonation of 32 mm^2, i.e. an effective r_g of 3.2 mm. It is likely that the energy loss factors in the vibrating glottis are also larger. These could be considered by increasing the visco-thermal attenuation coefficient in the glottal cylinder. A more realistic model of the glottal impedance, based on the work of Van den Berg (1957) and Stevens (Stevens, 2000, p. 165), includes a real impedance with both a resistive term relating to the viscosity, and a turbulence term related to the volume flow

$$Z_g = \left[\frac{12\mu h_g}{l_g d_g^3} + 0.875\frac{\rho U_g}{2 l_g^2 d_g^2}\right] + j\frac{\rho\omega h_g}{l_g d_g} \quad (4\text{-}1)$$

where U_g is the volume flow through the glottis, the coefficient of air viscosity $\mu = 1.8\cdot10^{-5}$ kgm^{-1}s^{-1}, glottal thickness $h_g = 3.2$ mm, and $l_g = 18$ mm (Van den Berg *et al.*, 1957), and the glottal diameter $d_g = 2r_g$.

However, for the very simple model calculation presented here, the two methods of determining Z_g (from equation (2-2) and (4-1)) yield sufficiently similar results, so the simpler small cylinder calculation is used, to avoid the need for additional parameters.

4.2.1.2.2. The impedance of the subglottal tract

Lulich *et al.* (2011) report that it is reasonable to consider the subglottal tract as an open cylinder of length $180 \leqslant l_{sg} \leqslant 235$ mm with a flanged opening into a radiation field. This model satisfactorily predicts for the frequencies of the second and third subglottal resonances, but may underestimate the first by ~100 Hz. Dalmont *et al.* (2001) showed that the impedance of such a cylinder is approximately

$$Z_{sg} = -j\frac{\rho c}{\pi r_{sg}^2}\tan(\Gamma_{sg}(l_{sg} + \delta)) \quad (4\text{-}2)$$

where Γ_{sg} depends on the radius r_{sg} according to equation (2-3), and also on the additional effective length for a flanged opening

$$\delta = \left[0.8216 r_{sg}\left(1 + \frac{(0.77\Gamma_{sg} r_{sg})^2}{1+0.77\Gamma_{sg} r_{sg}}\right)^{-1}\right] \quad (4\text{-}3)$$

The lower frequency of the first subglottal resonance is presumably related to a combination of geometry and perhaps the rigidity of the walls (or possibly related to unknown transfer functions in the reported measurements), but this is not corrected in the simplified model of the present study.

4.2.1.3. Comparison of measurement and model impedance from the lips

The impedance measured at the lips can be calculated via recursion using equation (2-2), where the open subglottal tract cylinder is the load on the glottal cylinder, which in turn is the load on the non-rigid vocal tract cylinder. The resulting calculated impedance spectra at the lips are shown in Figure 4-3 for the closed glottis, i.e. $Z_g = \infty$, and for a glottal cylinder of r_g = 1 mm and l_g = 18 mm. These were chosen to give qualitative agreement between the calculated curve and the male impedance data in Table 3-1. In principle the Γ_g and Γ_{sg} may be affected by higher α than are derived for the vocal tract. However, in the absence of further information, here the impedance for each of the cylinders is calculated with the same wavenumber.

When the magnitude of Z_g is high, the glottal load on the vocal tract is inertive, and for low frequencies, this inertance appears in parallel with the tissue inertance. So the details of the subglottal tract impedance have negligible influence on the impedance spectra at the lips. The influence of the subglottal tract for the low values of Z_g appropriate to inhalation is discussed below in section 4.2.2.

At low frequencies, the tract is approximately a compact object so, in this simple model, the high glottal inertance appears in parallel with the tissue inertance. Consequently, the total inertance in parallel with the compliance of the air is reduced by opening the glottis, which thus increases the f_{A0} (see logarithmic axis on Figure 4-3), as observed (see Figure 4-1 and Figure 4-2). With a constant f_{R0}, one would also expect the magnitude and bandwidth of $A0$ to decrease with glottal opening, also as observed (see example measurement in Figure 3-6).

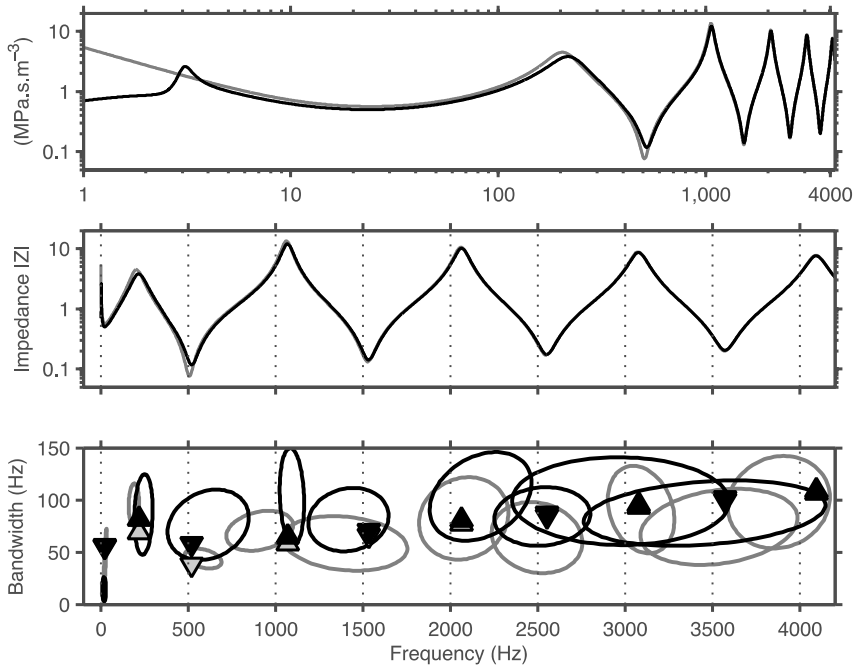


Figure 4-3 (top) Two models of the impedance of the vocal tract shown with a logarithmic frequency axis. Pale: closed glottis as shown in Figure 3-5. Dark: r_g = 1 mm, l_g = 18 mm. (Middle) The same curves with a linear frequency axis. (bottom) Bandwidth *B* for the closed glottis measurements (pale ellipses), for the model shown above (pale triangles) and for the phonation measurements (dark ellipses) and its model (dark triangles).

In summary, for the acoustic resonances, *R*i and *A*i, phonation increases *B* while increasing $|Z|_{Ri}$ and decreasing $|Z|_{Ai}$. Some of this increase in *B* is accounted for by the finite glottal load Z_g but further losses are perhaps due to the DC flow increasing visco-thermal losses, and/or turbulence in the jet of air from the glottis. For the mechanical resonances, f_{A0} increases, and *B* and $|Z|$ of both *A*0 and *R*0 decrease.

4.2.1.4. The impedance at the vocal folds from the glottis

4.2.1.4.1. The impedance seen by the glottis

The next step is to use the measurements made at the lips and a vocal tract model to infer properties of the tract when excited at the glottis. 'Looking in' from the lips, the vocal tract looks like a duct closed or almost closed at the glottis. 'Looking out' from the glottis, the vocal tract looks like a duct open at the lips. Thus, at lowest order approximation, the resonances (impedance minima) measured at the lips are anti-resonances (impedance maxima) at a closed glottis, and *vice versa*.

For low frequencies and a compact vocal tract, the parameters associated with the mechanical resonances of the tract do not depend upon which end of the tract is used for measurements. For higher frequencies, and provided that the simple cylindrical model is retained, the values estimated for the attenuation constant α would be the same for excitation at both ends. However, bandwidth measurements made with the subjects' lips sealed around the impedance head could differ from bandwidths that would apply for excitation at the glottis when the subject has an open mouth. Thus the bandwidth measurements of the resonances in the preceding section are not directly comparable with formant bandwidths reported in the literature. For example, the transmission loss due to the resistive term of the radiation impedance of an open cylinder causes the magnitudes of the impedance extrema calculated at the glottis to decrease more rapidly with frequency than those of a cylinder closed at the lips.

So, to consider the impedance at the glottis 'looking out' into the vocal tract: the tract is a non-rigid cylindrical duct, as modelled previously, but is open at the lips with a large flange (the face) as in equation (4-2). This impedance is an estimate of the impedance as 'seen' by the glottis looking toward the open mouth. In Figure 4-4 the impedance of the vocal tract is shown through glottal radii r_g = 7 mm and 1 mm. Decreasing the radius of the glottis (increasing the acoustic inertance) causes the extrema in the impedance spectrum to become asymmetric. Thus, for small r_g the minima in $|Z|$ decrease in frequency to approach the almost stationary maxima in $|Z|$ as described in Wolfe *et al.* (2009b).

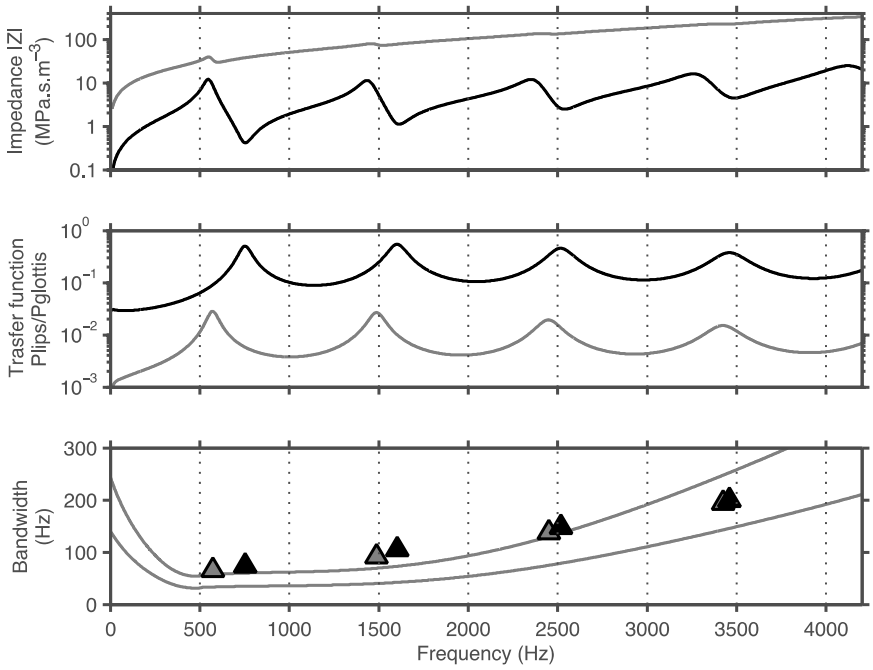


Figure 4-4 Calculated impedance magnitude of the vocal tract from the glottis (top), pressure transfer function ($P_{lip}/P_{glottis}$) (middle), and triangles showing the bandwidth of the transfer function maxima (bottom) for a vocal tract with glottal radius r_g = 5 mm (dark) and 1 mm (pale). The solid lines in the lower plot show two formant bandwidth estimations for f_0 = 80 Hz and 300 Hz using the estimation equation from Hawks and Miller (1995).

4.2.1.4.2. Transfer function estimation

The top graph in Figure 4-4 shows that, for a glottis with effective radius 1 mm, the inertance dominates the load impedance. For a radius of 5 mm, however, the resonances of the vocal tract dominate. It is expected that the maxima in the pressure transfer function from glottis to lips largely determine the formants *F*i. The frequencies and bandwidths of the maxima are shown as triangles in the lower plot of Figure 4-4. These B_{Fi} can then be directly compared to the values in the literature determined using external measurements. Hawks and Miller (1995) determined an empirical f_0 dependent relationship between f_{Fi} and B_{Fi} based on published closed glottis swept sine excitation measurements with a closed glottis (Fant, 1961; Fujimura and Lindqvist, 1971). This relationship is shown in Figure 4-4 along with the values of B_{Fi} determined in this fashion. In this context, one should note that estimates of B_{Fi} during phonation are greater than the estimates based on closed glottis data. For example, B_{Fi} determined from spectral examination of recorded utterances has been reported between 67 and 333 Hz, with median values of B_{F1}, B_{F2}, and B_{F3} of 130, 150

and 185 Hz respectively (Bogert, 1953). Pham Thi Ngoc and Badin (1994) reported an increase in B_{F1} of 40-120% from closed glottis to phonation for the French vowel [a] (where f_{F1} is close to the f_{R1} measured in this study) and up to 140% for higher resonances. The difference in the model values presented here is not so large since the extra losses at the glottis have not been incorporated into the model. Nevertheless, the differences in B_{Ri} for the model calculations of the transfer function show increases of 20-100% (see Figure 4-2).

Figure 4-4 also illustrates that increasing the glottal radius increases f_{F1}, which agrees with the experimental evidence on comparisons between phonated and whispered vowels (Jovicic, 1998; Kallail and Emanuel, 1984; Swerdlin *et al.*, 2010), the latter presumably having a larger effective r_g.

4.2.1.4.3. Implications for closed vowels

The value of B_{Fi} indicated in Figure 4-4 depends strongly on the boundary conditions of the tract at the lips. The impedances from which the parameters of the vocal tract model are derived are measured just inside the lips. So a more accurate model of the impedance at the glottis would include an additional 10 mm in length to account for the position of the measurement head inside the lips, plus the end effect to simulate radiation impedance. Furthermore, the vowel gesture used in this study, when spoken under normal conditions, would have an effective lip aperture smaller than that enforced by the experiment. In Figure 4-5 the pale curve shows the impedance and transfer function from the glottis where the effective lip radius is 7 mm. Compared to the 10 mm radius and shorter vocal tract in Figure 4-4 the determined values of B_{Fi} with the slight lip closure, shown with pale triangles in Figure 4-5, are slightly decreased, and occur at lower frequency due to the longer vocal tract.

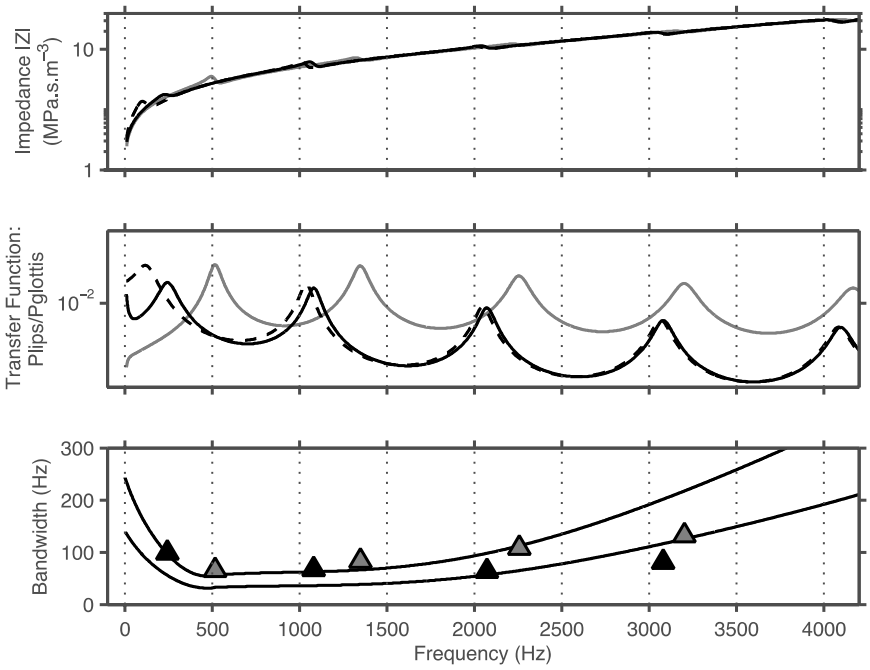


Figure 4-5 Calculated impedance magnitude at the glottis (top), pressure transfer function ($P_{lip}/P_{glottis}$) (middle), and bandwidth of the transfer function maxima (bottom) for a vocal tract with additional lip length of 10 mm (*cf.* no lip in Figure 4-4). The pale lines and triangles show the relatively open effective lip radius = 7 mm, and dark lines and triangles show an almost closed effective lip radius = 1 mm for the same non-rigid vocal tract model. The dashed lines show a rigid vocal tract with almost closed lips radius = 1 mm. The solid lines in the lower plot show formant bandwidth estimations for f_0 = 80 Hz and 300 Hz (Hawks and Miller, 1995).

Figure 4-5 also includes the impedance calculated for almost closed lips to demonstrate that B_{F1} increases as f_{F1} decreases below ~500 Hz, as observed by Fujimura and Lindqvist (1971). Hawks and Miller (1995) considered this behaviour by using two different sets of parameters for low and high frequency in their B_{F1} estimation. However, their calculation allows for unrealistically low f_{F1} since it does not consider how the non-rigid vocal tract influences f_{F1}. If the vocal tract were rigid, decreasing the effective lip radius would decrease f_{F1} until the impedance curve reached an infinitely high impedance at low frequency. However, the inertance of the non-rigid wall (discussed in 3.2.2) acts in parallel with the increasing inertance of the shrinking lip aperture, so the lower limit of f_{F1} is determined by the non-rigid wall anti-resonance at f_{A0}. Figure 4-5 shows examples of these two situations, with a dashed line for the rigid vocal tract and a solid dark line for the non-rigid vocal tract. The lower limit of f_{F1} occurs at ~300 Hz for men and ~320 Hz for women

using the average data from the vocal tract configurations in this study. It may be possible to make changes to the area of the tract, i.e. to deviate from cylindrical geometry, to lower this limit. However, it should not be possible for f_{F1} to occur lower than the closed glottis f_{A0} reported in Table 3-1 (210 and 245 Hz for men and women respectively). This finding also has implications to the acoustic load on the vocal folds with the large inertive loads used in straw phonation, which is discussed in Chapter 5.

Finally, the combination of a low f_{F1} due to lip closure and the mechanical properties of the wall may be important for the production of bilabial plosives such as [b], in which the initial broadband sound burst produced is observed to have a characteristic low frequency prominence and downward spectral tilt (Delebecque *et al.*, 2013; Kewley-Port, 1983; Stevens, 2000).

4.2.2.Low glottal impedance

The measurements made during low pitched phonation presented above do not show evidence of the influence of the subglottal tract, as the glottal aperture is small. Consequently, the glottal impedance is large and dominates the impedance 'looking down' from the glottis. Measurements during inhalation, when the glottal aperture approaches its maximum and the glottal impedance is reduced, are presented below.

4.2.2.1. Measurements during inhalation

A typical measurement made during inhalation is shown in Figure 4-6. The impedance spectrum shows several additional extrema when compared to the impedances measured with glottis closed in Figure 3-4 and during phonation in Figure 4-1.

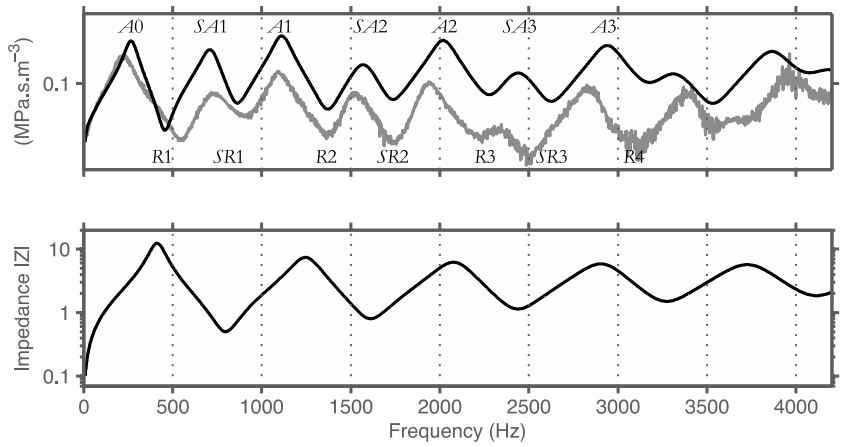


Figure 4-6 (top) The pale line shows the measured impedance magnitude spectrum at the lips during inhalation for one male subject (examples for each subject are given in Appendix A). The maxima and minima are labelled as in Table 4-1 and Table 4-2. The dark lines show calculated impedance magnitude at the lips (top), calculated impedance magnitude at the glottis looking into the lungs (bottom) for a vocal tract with glottal radius r_g = 7 mm, and length l_g = 18 mm; and, subglottal tract length l_{sg} = 180 mm and subglottal tract radius r_{sg} = 8 mm.

Since the closed glottis and phonation data in Table 3-1 show the positions of the resonances and anti-resonances of the vocal tract, one can infer which resonances are associated with the vocal tract and which with the subglottal tract. Table 4-1 and Table 4-2 summarise the mean measured data for men and women respectively, where *SR*i are the resonances and *SA*i the anti-resonances seen at the mouth due to the acoustic coupling to the subglottal tract.

Table 4-1 Mean male data from impedance measurements during inhalation.

Male Resonance	*R*1	*SR*1	*R*2	*SR*2	*R*3	*SR*3	*R*4
Frequency (Hz)	530±60	880±55	1335±145	1735±70	2210±75	2565±95	3660±270
$\|Z\|$ (MPa·s·m^{-3})	0.25±0.17	0.75±0.85	1.2±1.6	1.2±1.8	0.35±0.30	0.48±0.43	0.39±0.41
B (Hz)	45±55	90±85	70±65	50±60	105±80	65±60	50±40
No. samples	281	261	252	159	127	209	173
Anti-resonance	*A*0	*SA*1	*A*1	*SA*2	*A*2	*SA*3	*A*3
Frequency (Hz)	240±50	725±70	1090±85	1530±65	1930±95	2315±73	3080±185
$\|Z\|$ (MPa·s·m^{-3})	4.8±4.9	1.5±1.6	2.9±4.9	2.2±2.7	7.0±9.2	2.0±2.8	8.4±6.9
B (Hz)	55±45	75±70	55±55	110±100	40±35	100±80	50±35
No. samples	319	263	243	201	162	147	181

Table 4-2 Mean female data from impedance measurements during inhalation.

Female Resonance	*R*1	*SR*1	*R*2	*SR*2	*R*3	*SR*3	*R*4
Frequency (Hz)	660±30	1020±45	1490±45	1820±65	2460±55	2790±160	3680±455
$\|Z\|$ (MPa·s·m^{-3})	0.28±0.12	0.20±0.01	0.13±0.13	0.18±0.01	0.35±0.31	0.30±0.22	0.39±0.16
B (Hz)	70±65	25±15	15±10	70±80	85±85	65±55	80±75
No. samples	103	32	14	63	34	67	33
Anti-resonance	*A*0	*SA*1	*A*1	*SA*2	*A*2	*SA*3	*A*3
Frequency (Hz)	260±35	785±60	1200±60	1680±70	2170±60	2505±180	3480±205
$\|Z\|$ (MPa·s·m^{-3})	7.6±6.0	0.68±0.15	1.1±1.2	0.48±0.37	1.3±0.52	1.8±1.7	2.4±2.8
B (Hz)	65±30	70±55	75±70	75±80	85±55	85±65	70±50
No. samples	182	117	103	104	67	112	81

There is little scatter in the measured frequency of the additional resonances. However, since the prominence of the resonances depends on the glottal aperture, which was difficult to keep constant, there is a large range of $|Z|$ and B for these resonances.

Reasonable agreement with the measured impedance spectra in Figure 4-6 is obtained with a model of the subglottal tract impedance from equation (4-2) with length l_{sg} = 180 mm and radius r_{sg} = 8 mm. However, since this is an inversion problem and there are numerous parameters that can influence the shape of the impedance spectrum it is worthwhile to briefly summarise the effect of each parameter:

- l_{sg}: increasing the length of the subglottal tract decreases f_{SRi} and f_{SAi}.
- r_{sg}: decreasing the radius increases the characteristic impedance of the subglottal tract and increases the prominence of the measured peaks.
- $\alpha_{sg:}$ increasing the attenuation coefficient decreases the prominence.
- $\alpha_{g:}$ increasing the attenuation coefficient of the glottis also decreases the prominence.
- r_g: increasing the glottal radius increases the prominence of the peaks and increases f_{SRi} and f_{SAi}.

With this simple Z_{sg} model Figure 4-6 shows that f_{SAi} correspond to the minima in the impedance as seen by the glottis looking into the subglottal tract. However, as shown earlier, the inertance of the glottal aperture increases the asymmetry of the impedance seen though the glottis. So the impedance minima seen from the glottis when r_g is smaller than r_{sg} are lower in frequency than those that would be measured with no glottis, e.g. the swept sine measurements on cadaveric specimens (van den Berg, 1960). Therefore the f_{SAi} in Table 4-1 and Table 4-2 can be considered as a lower frequency limit on the minima of the subglottal impedance.

The maxima of Z_{sg} do not change frequency significantly when viewed through the glottis. However, they do not correspond directly to the f_{Sri} or f_{Sai} measured through the lips. Lulich *et al.* (2011) made measurements of the vibration of the skin below the glottis through the unknown transfer function of the neck. They reported vibrational maxima at 400-900 Hz, 1100-1700 Hz and 1800-2600 Hz for 25 male and female subjects, and assumed that these correspond to the maxima in the subglottal impedance. Both the values of f_{SAi} and f_{SRi} in Table 4-1 and Table 4-2 are within the range of those measurements. However, inferring the exact frequencies of the extrema of Z_{sg} from the values of Lulich *et al.* is highly dependent on the model of the glottal source and the relevant transfer functions.

4.2.2.1. Contribution of subglottal impedance

For a sufficiently high glottal impedance (small effective glottal radius), the subglottal tract does not affect the impedance measured through the lips. As the glottal impedance decreases, with radius increasing from approximately $r_g > 1$ mm, the increasing coupling to the vocal tract causes the resonances associated with the subglottal tract to introduce additional extrema into the impedance measured at the lips. This is in agreement with similar modelling by Fant *et al.* (1972).

Larger glottal areas have been reported during low pitched phonation. Hoppe *et al.* (2003) reported a maximum glottal area equivalent to $r_g = 3$ mm For such a large aperture the subglottal impedance is expected to contribute to the impedance at the mouth. For comparison, a calculation with $r_g = 3$ mm is also shown in Figure 4-7. In this case an additional impedance peak is visible close to the first resonance frequency of the subglottal system. This feature was not present in the phonation data presented here, since the average r_g over the time window of the measurement is presumably much lower than this value and/or the loss in the glottis is increased (e.g. increased α) with respect to that in the rest of the tract.

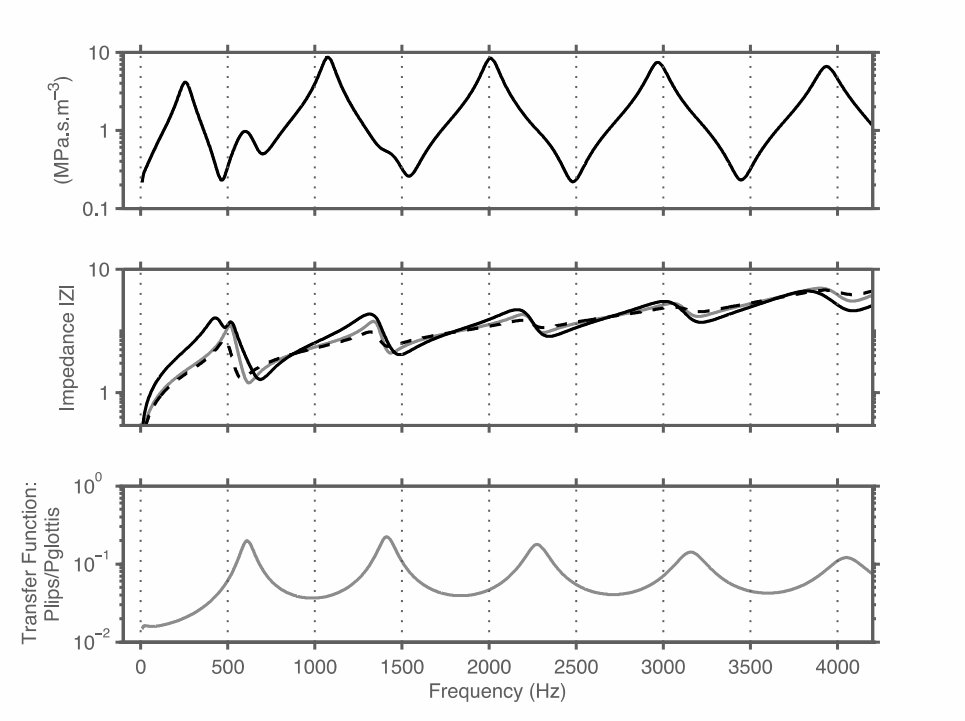


Figure 4-7 Calculations of impedance magnitude at the lips (top), impedance magnitude at the glottis (middle), and pressure transfer function ($P_{lip}/P_{glottis}$) (bottom) for a large glottal radius r_g = 3 mm. This is comparable to the maximum observed during low pitched phonation (Hoppe *et al.*, 2003). The subglottal tract parameters are the same as those in Figure 4-6. The solid pale line shows the impedance calculated independently of the subglottal tract. The solid dark line shows the impedance of the glottis considered as loaded with the series impedances of Z_{vt} and Z_{sg}. The dashed line shows the glottis with these impedances providing the load in parallel. In both cases, an extra peak in the transfer function occurs close to the minimum in Z_{sg} (see Figure 4-6). The maxima in Z_{sg} do not affect the transfer functions.

In Figure 4-4 the impedance as seen by the glottis and the pressure transfer functions with two different glottal apertures are displayed. The upstream impedance of the subglottal tract in these cases is neglected, since for impedance calculations, the acoustic load upstream of the reference measurement plane is unimportant. However, the vocal folds are likely to experience some combination of the impedance from upstream and downstream.

4.2.2.2. Estimation of driving impedance at the vocal folds

It is not straightforward to consider the acoustic load acting on the vocal folds. Vertical motion of the vocal folds, something like an outward swinging door (Fletcher, 1993; Helmholtz and Ellis, 1875), would be driven by the pressure difference $P_1 - P_2$ across it. From continuity, $U_1 = -U_2$, and $Z_1 = P_1/U_1$ and P_2/U_2, it follows that $P_1 - P_2 = U_1(Z_1 + Z_2)$: the series impedance of the subglottal tract and vocal tract (Benade,

1985). However, it is also possible that some of the force separating the vocal folds is provided by the average pressure in the glottis, like the sliding door motion described by Helmholtz and Fletcher, so the vocal folds might be affected by Z_{sg} and Z_{vt} in parallel, as suggested by Stevens (2000), Chi and Sonderegger (2007), and Lulich (2010).

Figure 4-7 shows the impedance and transfer functions from the glottis loaded by either: Z_{vt} (pale), the series addition of Z_{sg} and Z_{vt} (dark); and the parallel addition of Z_{sg} and Z_{vt} (dashed). Both the series and parallel load impedances show additional peaks associated with the minima of the Z_{sg}. Such features have been observed as formants in the spectra of *in vivo* phonation (Fant *et al.*, 1972) and swept sine measurements with excitation close to the open glottis *in vivo* (Fujimura and Lindqvist, 1971). These spectral features may influence the characteristics of breathy phonation, where the average r_g is large.

In Figure 4-4 and Figure 4-7, the impedance of the vocal tract is considered as seen from the lower part of the glottis, i.e. through its entire length. Since flow separation is expected to occur at the upper part of the glottis, it may be more useful to consider the impedance at the top of the glottis, i.e. looking through the end effect δ from equation (4-3), as the impedance seen by the vocal folds. Figure 4-8 is comparable with Figure 4-7 except that only the end effect of the glottal length is considered. This has a very similar effect on the impedance curve to increasing the glottal radius, as shown in Figure 4-4, i.e. the impedance extrema become more symmetric, and the frequencies of the pressure function maxima increase.

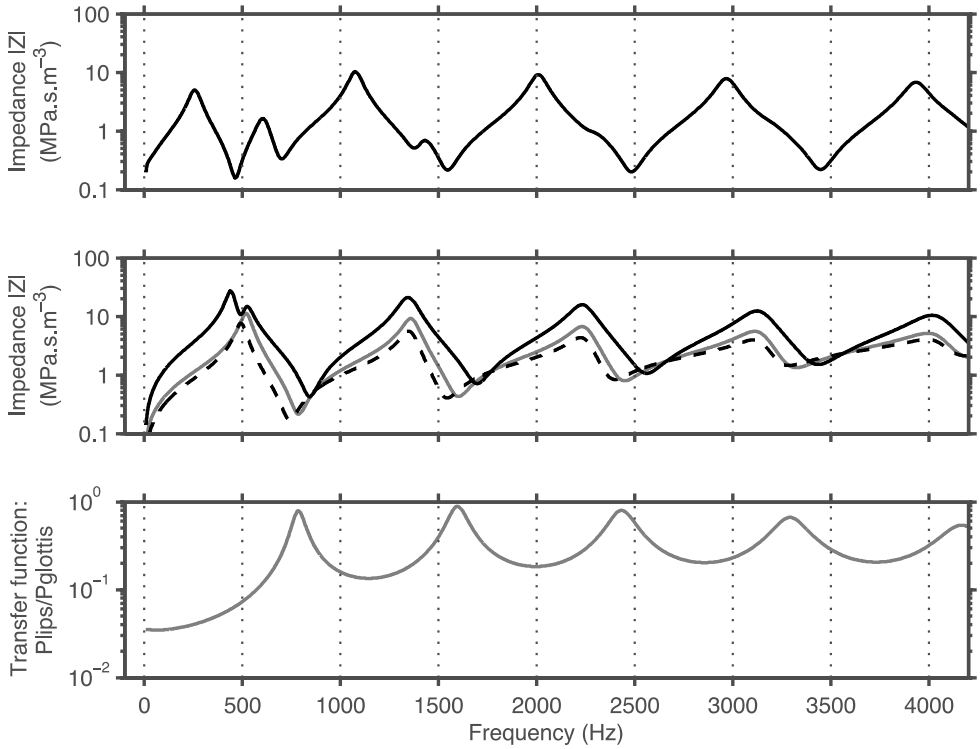


Figure 4-8 Calculations of impedance magnitude at the lips (top), impedance magnitude through the glottal end effect (middle), and pressure transfer function ($P_{lip}/P_{glottis}$) (bottom) for a large glottal radius r_g = 3 mm. *c.f.* Figure 4-7. The solid pale line shows the impedance calculated independently of the subglottal tract. The solid dark line shows the impedance of the glottis considered as loaded with the series impedances of Z_{vt} and Z_{sg}. The dashed line shows the glottis with these impedances providing the load in parallel. In both cases, an extra peak in the transfer function occurs close to the minimum in Z_{sg} (see Figure 4-6). The maxima in Z_{sg} do not affect the transfer functions.

4.3. Summary

Low pitched, M1 phonation increases the bandwidths of the acoustic vocal tract resonances, but does not significantly change the resonance and formant frequencies, compared to closed glottis measurements. Some of this increase is accounted for by the inertive acoustic load that a partially open glottis adds to the vocal tract but it is expected that higher visco-thermal and/or turbulent losses due to the airflow through the glottis are also involved.

Considering the impedance of the glottis, two different behaviours are observed with high and low magnitude impedance:

For high glottal impedance (r_g < 1 mm): f_{A0} increases as Z_g is lowered. The frequencies of the acoustic resonances Ri do not change significantly at the 1% level, and Ai change little. The transfer function is highly asymmetric so that f_{Fi} and f_{Ri} of the measured impedance at the lips are close to the maximum in the output impedance at the glottis.

Resonances of the subglottal tract are not evident in the impedance measured through the lips during low pitched, modal phonation due to the small average glottal apertures (and possibly additional glottal energy loss factors).

For vowel production in speech and singing, $R0$ can be neglected since it lies below achievable f_0 even with period doubling phonation. However, $A0$ may influence the relative prominence of f_0 or its harmonics particularly when the mouth opening is small. Furthermore, f_{A0} imposes a lower limit to f_{F1}, since at low frequencies the wall inertance is in series with the inertance of the mouth opening. So it is not possible to reduce f_{F1} below f_{A0}. This also causes B_{F1} to increase as f_{F1} approaches this value in agreement with the literature. This is a possible reason for the limit of the vowel plane in the low $F1$ direction for all languages, which is an interesting observation about speech in general.

For low glottal impedance (r_g > 1 mm): the impedance at the glottis looking out becomes more symmetric, so that f_{Fi} and B_{Fi} increase, as observed when changing from normal phonation to whispering in both the resonances and the formants.

Since some types of vocal fold motion may be driven by the series combination of Z_{vt} and Z_{sg} and other types possibly by the parallel combination, depending on the model, pressure transfer functions under these two conditions were calculated and both show peaks close to the minima of Z_{sg}, which may be observable in breathy phonation.

5 *Ex Vivo*: Phonatory Dynamics of Excised Larynges

This chapter explores the P_{sg}, f_0 relationship of human excised larynges as a function of the type of laryngeal control[7]. The vibratory behaviour was controlled aerodynamically, by adjusting the air supply, and mechanically, by adjusting the laryngeal geometry corresponding to a combination of laryngeal muscle movements. No vocal tract was present in the experiments, so the acoustic load downstream of the vocal folds is small and inertive. In this configuration the impedance of the subglottal tract may be expected to influence the motion of the vocal folds, so the upstream impedance of the air supply was estimated and its influence on the vocal folds investigated.

5.1. Material and Methods

5.1.1. Larynx preparation

This experiment was carried out on cadavers donated to the Anatomy Laboratory of the French Alps (Laboratiore d'Anatomie des Alpes Françaises). Twelve larynges (six from men and six from women) were dissected within 48 hours post-mortem. The larynges were kept intact from the epiglottis to the fifth tracheal ring. None of the epilaryngeal structures were removed. An X-ray computed tomography (CT) scan made of one larynx dissected in this way was used to make a 3D visualisation of its cartilage, shown in Figure 5-1 (Chiche, 2012).

[7] I acquired the data for the experiments in this chapter, except where otherwise noted in the acknowledgements. I performed all of the data analysis for this chapter under the supervision of Nathalie Henrich Bernardoni. Significant portions of this chapter have been submitted for publication as 'Laryngeal control of fundamental frequency in excised human larynges' with co-authors N. Henrich Bernardoni, F. Bennani, L. Morand, M. Pelloux, A. Mancini, X. Laval, T. Legou, P. Chaffanjon (Hanna *et al.*, 2014a). A pilot study based on data from a single larynx (LF4) was presented in 'Singing excised human larynges: relationship between subglottal pressure and fundamental frequency,' International Workshop on Models Analysis of Vocal Emissions for Biomedical Analysis (MAVEBA), Florence, Italy (Hanna *et al.*, 2013a), and further preliminary portions of this work were presented in 'Singing excised human larynges: Investigating aerodynamical and biomechanical control of phonation,' International Conference on Voice Physiology and Biomechanics (ICVPB), Salt Lake City, Utah, USA (Hanna *et al.*, 2014b). The design of the experimental equipment was reported as 'Construction d'un modèle laryngé hybride. Mise au point d'un banc d'essai' (Robert *et al.*, 2014).

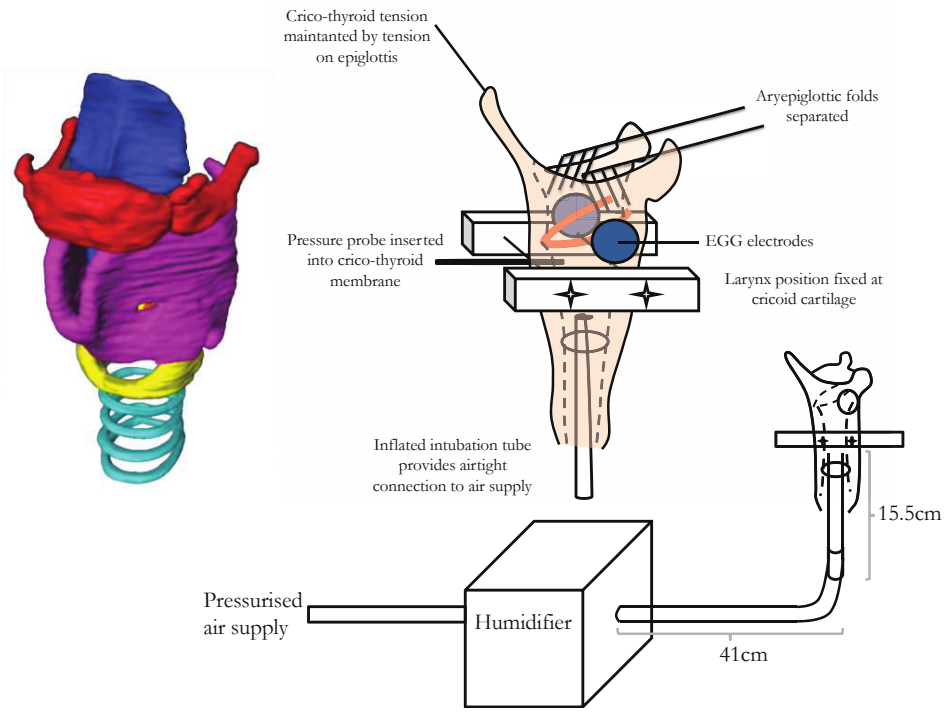


Figure 5-1 (left) 3D visualisation of the cartilage of one of the dissected larynges from a CT scan, reproduced with permission from Chiche (2012). The epiglottis is coloured blue, the hyoid bone is red, the thyroid cartilage is purple, the cricoid cartilage is yellow, and the tracheal rings are yellow and aqua. (middle right) Illustration of a larynx attached to the experimental bench by the cricoid cartilage. (lower right) A sketch of the the upstream air supply system.

Details of the larynges are given in Table 5-1. For each experiment, the larynx was secured to the experimental bench posteriorly and anteriorly at the cricoid cartilage (shown in Figure 5-1). An intubation tube was used to supply a source of heated humidified air from a pressurised reservoir (Mecafer, LT50 Compressor) with adjustable constriction. Crico-thyroid tension was maintained by securing the epiglottis. The aryepiglottic folds were separated by the use of a Beckman-Eaton laminectomy retractor. To examine the behaviour of the supraglottal larynx, additional measurements were made without this separation.

Table 5-1 Description of the excised larynges tested, including the recorded f_0 ranges in hertz and musical notation (in square brackets). Women's larynges are denoted LF, and men's larynges are denoted LM. The modes correspond to two distinct regimes in the P_{sg}, f_0 plane, following the terminology of Alipour and Scherer (2007).

Larynx	Condition	Age (years)	Height (m)	Weight (kg)	Air supply control				Mechanical control			
					1st Mode		2nd Mode		1st Mode		2nd Mode	
					Min	Max	Min	Max	Min	Max	Min	Max
LF1	fresh	96	1.50	40	83 [E1]	226 [A3]	192 [G2]	664 [E4]			219 [A3]	467 [A#4]
LF2	frozen	66	1.65	66	70 [C#1]	235 [A#3]	238 [A#3]	931 [A#5]	95 [F#2]	241 [B3]	318 [D#4]	746 [F#5]
LF3	frozen	96	1.50	45	224 [A3]	340 [F4]			211 [G#3]	690 [F5]	215 [G#3]	687 [F5]
LF4	frozen	81	1.65	65	65 [C2]	895 [A5]			292 [D4]	713 [F5]		
LF5	frozen	80	1.55	40								
LF6	frozen	83	1.80	85								
LM1	fresh	72	1.80	75	67 [C2]	204 [G#3]			175 [F3]	520 [C5]	253 [B3]	510 [C5]
LM2	frozen	75	1.70	75	130 [C3]	145 [D3]	100 [G#2]	196 [G3]	120 [B2]	495 [B4]	140 [C#3]	624 [D#5]
LM3	frozen	69	1.75	70	66 [C2]	229 [A#3]	87 [F2]	137 [C#3]	102 [G#2]	382 [G4]	123 [B2]	339 [E4]
LM4	frozen	87	1.80	70	65 [C2]	207 [G#3]	65 [C2]	207 [G#3]	110 [A2]	423 [G#4]	110 [A2]	539 [C5]
LM5	frozen	79	1.70	80								
LM6	fresh	83	1.60	60								

Dehydration of the larynx was observed by colour changes in the tissue and eventually as an increase in stiffness, which can result in tears and holes in the soft tissue. The larynges were disposed of appropriately once they were no longer able to sustain self-oscillation.

No acoustic loading corresponding to a vocal tract was included in the experiment. Therefore the downstream acoustic load on the vocal folds is small and inertive, which is expected to be beneficial to vocal fold oscillation according to one model (Titze, 2008). An

artificial vocal tract would provide a more realistic setting but would also introduce similar constraints on access to the larynx as an *in vivo* study and therefore remove some of the advantage of an *ex vivo* approach.

With no supraglottal vocal tract, the influence of the subglottal impedance may be expected to cause bifurcations in the P_{sg}, f_0 oscillation (Zhang *et al.*, 2006). The impedance of the experimental setup used was estimated using the geometry of the plumbing upstream of the vocal folds, in a similar way to 4.2.1.4. The intubation tube connected to the larynx is 15.5 cm (diameter 1.6 cm), and is connected a larger diameter tube (2.1 cm) of length 41 cm that is terminated by several litres of air in the humidifier. A sketch of the plumbing is included in Figure 5-1.

5.1.2.Data acquisition

Data were acquired at a sampling rate of 6.25 kHz using a specialised aerodynamic workstation (EVA2, S.Q.Lab, Aix-en-Provence, France), then analysed using Matlab. The measurement and analysis processes for each variable are described below.

P_{sg} was measured with a probe inserted via a catheter into the crico-thyroid membrane. AC (audio) component of the signal was removed by low pass filtering (2^{nd} order Butterworth filter at 5 Hz) to provide the DC (aerodynamic) pressure.

The vocal fold vibratory behaviour was monitored by an electroglottograph (EGG) (EG2-PCX2, Glottal Enterprises, Syracuse, New York, USA). f_0 was estimated from the EGG signal using the YIN algorithm (de Cheveigné and Kawahara, 2002), with a frequency detection range of 65-1000 Hz and the absolute threshold parameter set to 0.5. Periods of aphonia and erratic quasi-periodic behaviour were removed from further analysis.

The sound pressure level was measured by the EVA2 workstation microphone at a distance of 30 cm from the larynx. The background noise level was ~45 dB, so the lower range of phonation sound pressure level was not recorded with sufficient signal to noise ratio. Consequently the reported dynamic ranges of each larynx are conservative.

Each experimental sequence was recorded with a Sony HVR-V1 HD video camera at 60 interlaced frames per second with a vertical resolution of 1920x1080 pixels. A Nikon EOSENS Cube 7 high speed camera was used to record a few sequences at 2000 frames per second for two of the larynges, LF1 and LM1, in order to assess vocal fold closure.

5.1.3.Experimental protocol

The study was conducted with two sweeping (increasing then decreasing) control conditions: air supply and mechanical. Saline solution was applied before each sweep.

For the air supply control condition, air from the pressurised reservoir was released by varying the output constriction, allowing air to flow through the humidifier and on to the larynx. The pressure P_{sg}, measured just below the glottis depends on the load at the glottis, and so is not constant on the time scale of the vocal fold oscillations. P_{sg}, was increased then decreased by hand, while the laryngeal geometry was kept fixed by applying one of three different applied tensions to produce a low, medium or high level of vocal fold adduction.

For the mechanical control condition, the air supply constriction was fixed to a value that allowed phonation with a minimum tension applied to the vocal folds. Two methods of simulating *in vivo* biomechanical laryngeal control were tested. For the majority of larynges, sutures were secured in the muscular processes of the arytenoid cartilages and pulled as illustrated in Figure 5-2, similar to the approach of Berry *et al.* (1996). For larynges LF4, LM5 and LM6, adjustment of vocal folds geometry was performed manually with metal probes inserted into the arytenoid cartilage, also shown in Figure 5-2.

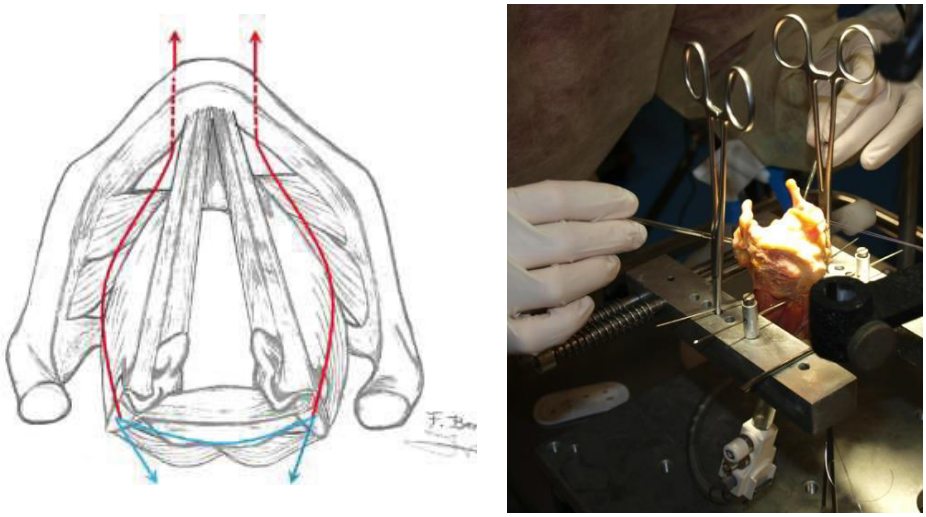


Figure 5-2 (left) A schematic view from above showing the sutures used to simulate vocal fold tension adjustments. (right) A photograph showing the adjustment of vocal fold tension with metal probes inserted in the arytenoid cartilage.

The aim of the mechanical control condition is to mimic a combined action of lateral crico-arytenoid and thyro-arytenoid muscular tension. The direction of applied tension was adjusted to achieve the most complete vocal fold adduction during oscillation, an example is shown in Figure 5-3. Tension was applied symmetrically by pulling the sutures from the same support via a dynamometer, then releasing. The applied tension was increased to a maximum of 10 N depending on the larynx. This approach adjusts the vocal fold geometry but does not directly activate the muscles, and since it is not possible to measure their internal stress, the effect may not be representative of the muscular laryngeal control performed *in vivo*. Furthermore, increasing the vocal fold adduction in this way increases P_{sg} and may cause other effects, e.g. adducting the vocal folds medially and therefore changing their effective mass.

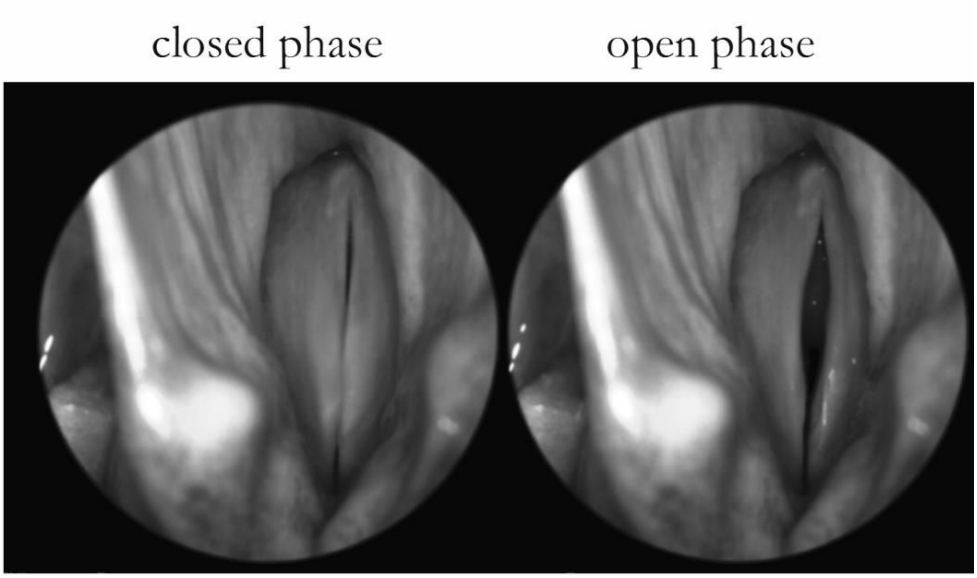


Figure 5-3 Example still images extracted from the 2000 fps video, showing the closed and open phases of vocal fold oscillation in larynx LM1 with the laryngeal configuration controlled by the sutures shown in Figure 5-2.

5.2. Results and Discussion

Firstly, the range of phonation achieved by each larynx is presented, and the limits of repeatability determined. Then the dependence of f_0 on air supply and vocal fold geometry is investigated, looking for qualitatively different regimes and quantifying the variation in each regime. An empirical power law for the P_{sg}, f_0 relationship is compared with a linear fit. Finally, the passive behaviour of aryepiglottic larynx is examined.

5.2.1.Assessment of *ex vivo* laryngeal oscillation

Table 5-1 summarises the observed range of fundamental frequencies for each larynx. Where two modes of phonation (i.e. two distinct regimes in the P_{sg}, f_0 plane) were observed, they are listed separately and are discussed in 5.2.3). Larynges LF5, LF6, LM5 and LM6 did not sustain vocal fold oscillation under any of the experimental conditions.

Differences in the behaviour of men's and women's larynges are to be expected. For example, the men's larynges have a lower possible f_0 due to the length and mass of the vocal folds (which controls the lowest mode of oscillation available). They also exhibit a lower maximum f_0, 624 Hz (D#5) *cf.* 931 Hz (A#5) for a woman's larynx in the measurements reported here. The women's larynges phonated at much higher P_{sg} than the men's. For larynges LF2 and LF3, increased P_{sg} did not result in higher sound pressure level than for the men's larynges.

The upper limit of the aerodynamic pressure probe was 10 kPa. In all but one case, phonation had ceased before this level was reached. The exception was a small woman's larynx LF1, and these measurement portions were excluded from results shown. Typically, phonation ceased when P_{sg} approached 6 kPa for both sexes.

The larynges in this study were dissected from cadavers donated for research. Such donations are necessarily anonymous and therefore individual medical histories are unknown. The cadavers are transported within 48 hours post-mortem, so may have been refrigerated for up to 48 hours before dissection. Despite such limitations, eight out of the twelve larynges tested were able to sustain self-oscillation under the air supply and mechanical control conditions, and over an extensive range of f_0 and sound pressure level. Measurements made directly after dissection on LF1 produced a greater dynamic range.

Mounting sensors, particularly the EGG electrodes, was problematic for the smaller female larynges. In these cases, the lack of good contact gave a weak signal and made f_0 estimation less accurate.

Estimation of f_0 is difficult for complex signals characterised by glottal-cycle perturbations. The YIN algorithm (de Cheveigné and Kawahara, 2002) was used with a minimum

frequency of 65 Hz. Some estimates of the larynx f_0 occur at this frequency, which may reflect aperiodicity in the signal.

It has been reported that slow freezing of canine vocal folds decreases the elastic shear modulus and dynamic range of the mucosa, so it is recommended for excised larynges to be cryopreserved in liquid nitrogen and warmed in saline solution overnight (Chan and Titze, 2003).

One frozen male larynx (LM4) was tested three times during a one-week period in which the larynx was frozen and thawed between each measurement session. Figure 5-4 displays the measurements for three successive sessions. The two modes of phonation and the dynamic range were retained throughout, however the f_0 range reduced by the third test. Each measurement session lasted approximately two hours. Despite the frequent reapplication of saline solution and the humidity of the subglottal air supply, the superior tissue of the larynx gradually dried and became stiffer during the measurement sessions, causing a decrease in the range of f_0 by the end of the second test. This suggests that freezing and thawing may cause less tissue damage than the dehydration during the measurement session.

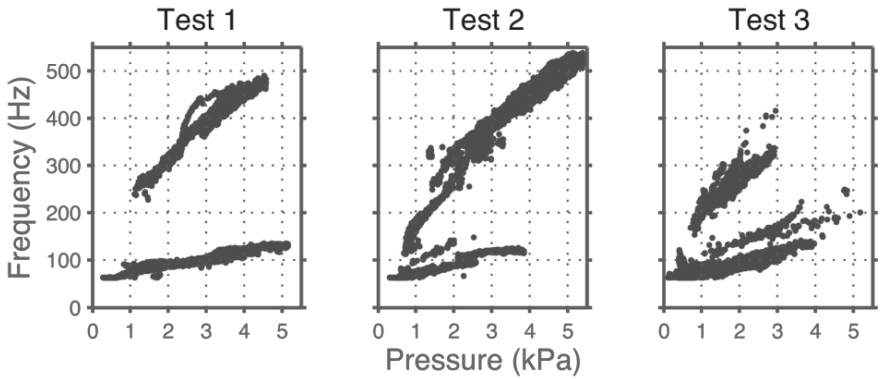


Figure 5-4 Larynx LM4 tested after three freeze/thaw cycles. f_0 v. P_{sg} shows two distinct modes of vibration, which differ in terms of glottal contact area.

5.2.2.Discontinuities and hysteresis

Cycles in the P_{sg}, f_0 plane exhibit four types of discontinuities and hysteresis, which are illustrated by examples in Figure 5-5.

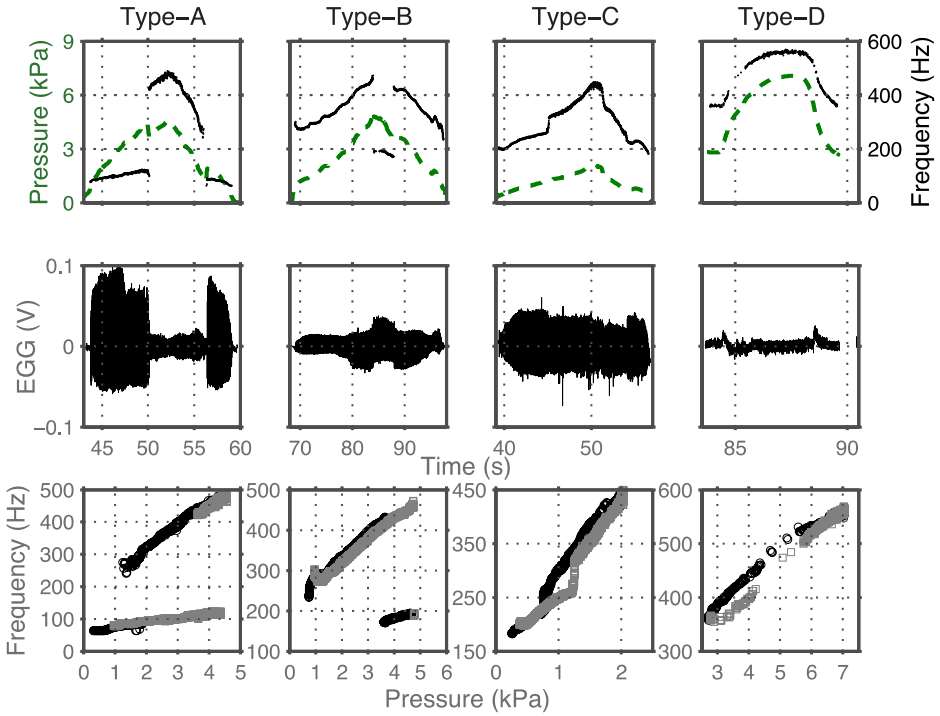


Figure 5-5 Examples of four types of hysteresis observed: A-D from left to right. The top plot shows f_0 (solid line) and P_{sg} (dashed line) with respect to time. The middle plot shows the EGG signal on the same arbitrary voltage scale with the same time axis as the top figures. The bottom P_{sg}, f_0 plots show the swept control parameter increasing (pale), and decreasing (dark). In these examples, type-A, B and C are demonstrated under air supply control, type-D under mechanical control. Note that P_{sg} increases due to the increased glottal resistance under mechanical control even though the air supply is fixed.

5.2.2.1. Type-A

This type of hysteresis was observed as a discontinuous increase in f_0 with increasing P_{sg} accompanied by reduced EGG contact amplitude above the jump, as illustrated in Figure 5-5. These differences in glottal contact recall the laryngeal mechanisms used in speech and singing to produce modal voice (M1) and falsetto voice (M2) (Roubeau *et al.*, 2009). None of the female larynges demonstrated such a pitch jump, whereas all male larynges demonstrated this type of behaviour at least once under the air supply control condition.

The only larynx with a large number of type-A pitch jumps recorded was LM4, in which both air supply and tension sweeps reliably triggered this behaviour. Details of the frequency leap intervals for that larynx are given in Table 5-2, which shows that the minimum leap interval is smaller under the decreasing control condition.

Table 5-2 (top) Recorded number (#) and mean leap interval (in semitones) for LM4 under air supply and mechanical control parameters. (bottom) Number (#) and mean leap intervals (in semitones) for men's and women's larynges undergoing type-B discontinuities during increasing (Inc) and decreasing (Dec) pressure sweeps.

Type-A		Pressure Inc		Pressure Dec
	#	Mean	#	Mean
LM4	25	18 ± 6	25	-15 ± 5
		Tension Inc		Tension Dec
LM4	16	18 ± 6	15	-15 ± 3

Type-B		Pressure Inc		Pressure Dec
LF1	8	-16 ± 5	9	17 ± 3
LF2	5	-25 ± 3	5	23 ± 1
LF3	1	-12	1	11
Female	14	-19 ± 6	15	20 ± 5
LM1	3	-15 ± 3	1	12
LM2	13	-16 ± 3	10	16 ± 4
LM3	18	-19 ± 3	17	17 ± 6
Male	34	-17 ± 4	28	17 ± 5

5.2.2.2. Type-B

An obvious discontinuity in f_0 leading to a subharmonic pattern in the EGG signal. The overall vibratory cycle length is approximately preserved, with similar glottal contact area but the shape of each glottal cycle was slightly modified from one cycle to the next.

Table 5-2 summarises the observed pitch leap intervals in semitones for the larynges that displayed type-B behaviour,. Note that LF4 and LM4 did not display this type of hysteresis under the air supply control condition and the mechanical condition did not yield many of these discontinuities and so the data are not shown.

In type-B cases, the leap intervals were negative on increasing pressure, i.e. the larynx shifted from high to low pitch, and positive with decreasing pressure, which indicates a transition to a subharmonic vibratory pattern. The opposite behaviour is seen with decreasing pressure. The magnitudes of the frequency intervals were comparable for ascending and descending jumps.

5.2.2.3. Type-C

Rapid changes in pitch were observed with no modification in the glottal contact area, as assessed from the EGG signal. Such changes were typically of small amplitude (less than

10%) but in some cases reached up to 50% of the f_0 magnitude. Since there was no obvious discontinuity in these cases and the changes may approach the maximum rate of pitch change, it is difficult to quantify their occurrence without establishing a set criterion for the minimum rate of *df/dP*. This behaviour was observed under all experimental conditions for all larynges.

5.2.2.4. Type-D

Increases of either control parameter lead to smooth increases in f_0. When the control parameter is reduced, f_0 decreases smoothly but at a higher frequency for the same P_{sg} than on the increase. This is consistent with a reduction in vocal fold tension due to visco-elastic deformation of the tissues.

5.2.2.5. Summary and measurement of leap intervals

Observed rapid or discontinuous changes in f_0 (pitch jumps) ranged from a few semitones up to 30 semitones, a limit which corresponds musically to about 2 octaves and a half. These pitch jumps can be divided into three categories: A – with change in glottal contact area, B – with subharmonic structure, C – rapid change with no discontinuity and no change in glottal contact.

Type-A hysteresis events show a large change in glottal contact area on the EGG signal. This may indicate a change in laryngeal mechanism similar to that seen *in vivo* when changing from mechanism M1 to M2 (Roubeau *et al.*, 2009). Such behaviour may require a decoupling of the layered vocal fold structure, affecting the vibrating mass responsible for f_0. This coupling-decoupling is expected to be difficult to achieve *ex vivo* since the vocalis muscle is not directly manipulated. However, one male larynx (LM4) consistently produced this type of hysteresis under both experimental conditions. This allowed a rare opportunity to measure the uncontrolled leap interval for the human voice, independent of the automatic compensation observed in living subjects (Baer, 1979). One such measurement was previously made by Švec *et al.* (1999) while analysing a film (Van den Berg and Tan, 1959) of an excised human larynx.

Miller *et al.* (2002) attempted to determine a characteristic leap interval for the singing voice *in vivo*, noting that for the five male voices studied, a leap of approximately one octave (12 semitones) was typical, with bass-type voices having larger intervals (see also (Švec *et al.*, 1999)). For the male bass-type voice in the Miller study, leap intervals of 14-19 semitones were observed, with a typical leap from Eb3 to G4 (155 to 392 Hz or 16 semitones). This

range is comparable with our *ex vivo* measurements of LM4, where a mean leap interval of 18 semitones under increasing control parameters was recorded (Table 5-2), for leaps from an initial f_0 of 69-135 Hz to a final f_0 of 140-435 Hz.

There are several differences in experimental method between the Miller *et al.* study and this one, notably that only a single control parameter was adjusted in this *ex vivo* study. Although independent control of airflow is attempted in the *in vivo* study it is unlikely to be achieved due to automatic compensation effects. Švec *et al.* (1999) distinguish between a transient and a final (stable) leap interval, perhaps for this reason. As the current study uses swept control parameters, which cause glissandi, the final pitch did not stabilise but continued moving in the swept trajectory. Hence, the method of calculating leap intervals reported here is comparable to Švec's transient intervals, and the magnitude of the intervals is likely to be greater than intervals measured *in vivo*. These intervals may therefore place an upper limit on the achievable range of transitions between the two singing mechanisms.

Finally, the control parameters are different. Here the mechanical control parameters of the larynx were adjusted, but in the *in vivo* study, subjects were asked to increase airflow. However, the results presented here show that for both the mechanical and air supply control parameters the mean leap interval is the same. The *in vivo* study does not report the case of decreasing the control parameter to change from M2 to M1. In the present study, the mean leap interval is reduced from 18 (ascent) to 15 (descent) semitones under both experimental conditions.

For type-B behaviour, the frequency ratios of the leap intervals were often close to 3:1, 4:1 or 2:1. This kind of harmonic matching before and after the pitch jump confirms the transition to a subharmonic vibratory pattern and is similar to behaviour reported in the literature (Švec et al., 1996; Tokuda et al., 2007).

Type-C hysteresis was characterised by rapid f_0 variations of small frequency amplitude (up to 50% of f_0). Such pitch jumps may give information about the upper limit of smooth transitions in pitch. *In vivo* experiments have shown that singers can intentionally produce pitch changes of order 100 semitones per second (Johan Sundberg, 1973; Xu and Sun, 2002). However, since in this *ex vivo* study such pitch changes are unintended and are often very small in magnitude a strict criterion is necessary to establish when such a jump has occurred.

Type-D hysteresis was observed as a smooth change in f_0 on increasing and decreasing the control parameter. This does not involve any discontinuities and is an expected property of biological tissue, consistent with a reduction in vocal fold tension due to visco-elastic deformation.

Under both control parameters, similar numbers of type-C and D hysteresis events were observed. Type-B behaviour was not typically seen under the mechanical control sequences. These results emphasise the importance of P_{sg} on the control of f_0, in addition to the contribution of vocal fold biomechanics, in particular length and tissue stress (Titze, 2011).

It has been reported that the subglottal impedance can cause discontinuities in the behaviour of laryngeal models (Zhang *et al.*, 2006). The upstream impedance magnitude calculated following the method in 4.2.1.4 has maxima of order 100 MPa·s·m^{-3} at 170 Hz, 470 Hz and 740 Hz and minima of order 0.1 MPa·s·m^{-3} at 280 Hz, 620 Hz and 940 Hz. In Figure 5-6 this is shown as the load impedance on both large and small glottal apertures. The maxima in the series impedance are relatively fixed compared to the maxima of the intubation tube, but the minima decrease with decreasing glottal radius.

If the discontinuous behaviour of the larynges were influenced by the subglottal impedance of the system one would expect pitch jumps to occur close to the resonance frequencies of the impedance at the glottis. In this experiment, however, a large variance in initial and final f_0 with pitch jumps was observed, implying that the subglottal resonances do not have such an effect. As noted above, type-A hysteresis in LM4 was observed with an initial f_0 of 69-135 Hz and a final f_0 of 140-435 Hz. Furthermore, type-B hysteresis was observed over an even wider range with initial f_0 of 370-935 and 85-625 Hz and final f_0 of 65-235 Hz and 83-437 Hz for women's and men's larynges respectively. Therefore, the discontinuities displayed appear to be a biomechanical feature and not triggered by acoustic loading.

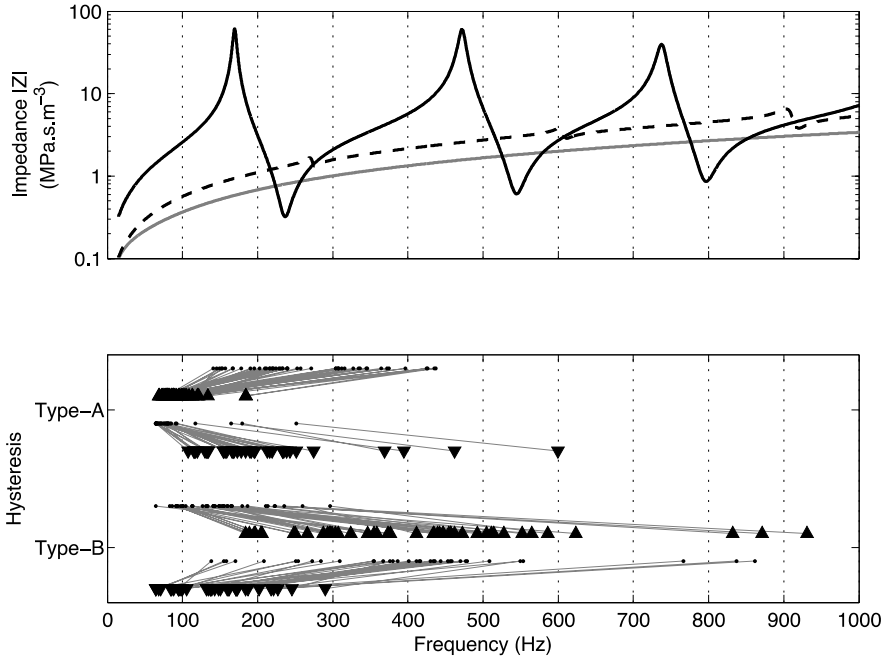


Figure 5-6 (top) Calculated impedance magnitude at the glottis, where the acoustic load is from the epilaryngeal larynx only (pale), from the epilaryngeal larynx in series with the upstream air supply plumbing (dark), and from the same two impedances in series (dashed). The impedance curves are calculated with a glottal radius r_g = 3 mm. (bottom) All type-A and type-B leap intervals. The initial f_0 are shown with upward pointing triangles for the increasing control parameter, and downward pointing triangles for the decreasing parameter. The overlapping ranges of the initial and final f_0 do not appear to be systematically related to the upper impedance curves, i.e. initial and final f_0 occur with both inertive and compliant loads.

5.2.3.Pressure-frequency relationship

Figure 5-7 displays all of the periodic portions of measurements under the two control conditions plotted on a P_{sg}, f_0 plane. Two distinct modes of vibration are visible, i.e. most larynges can vibrate at either a low or a high f_0 for a given value of P_{sg}. In the case of one larynx (LM4), the vibratory behaviour between these regions differs with respect to glottal contact area, as assessed by the amplitude of the EGG signal (see Figure 5-5). For the other larynges, there is no such change in glottal contact area. Following the terminology of Alipour and Scherer (2007) these distinct regimes are considered as the 1st and 2nd oscillation modes (see

Table 5-1).

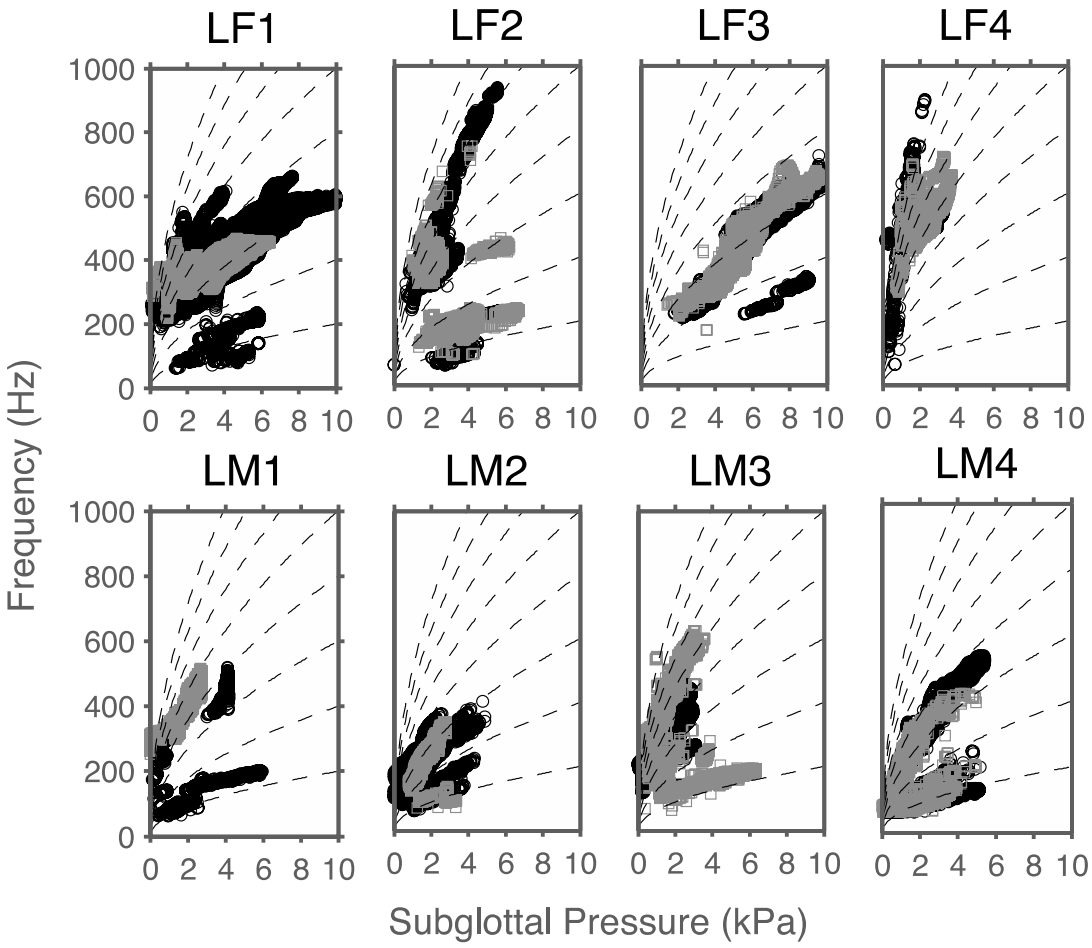


Figure 5-7 P_{sg}, f_0 relationship under air supply (dark circles) and mechanical (pale squares) control for all self-oscillating larynges. Dashed lines show the power law $f_0 = a \cdot P_{sg}^{\ b}$, with the mean value for all data of $b = 0.5$ and a varied from 5-50.

The calculations of the load impedance suggest that the experimental results include ranges of f_0 in both modes in which the larynges have both inertive and compliant vocal fold loads (the increasing and decreasing parts of the impedance magnitude curve shown in Figure 5-6). However, the glottal aperture is not measured for each larynx, so it is possible that the change from an inertive to a compliant load could occur in the region 180 to 300 Hz depending on the glottal aperture. Since, for some larynges, the upper limit of the 1st mode and the lower limit of the 2nd mode occur in this range, it is not possible to rule out the idea that the 1st mode occurs in the inertive region below the first subglottal resonance, and the 2nd mode in the inertive region above, with the compliant region in between separating the two modes.

Where there are two modes present the 2nd mode is considered to be the principal mode of vibration, as the 1st mode is often quasi-periodic, and may be made up of subharmonics of the 2nd mode. Discontinuities were removed from the phonation data and an unconstrained fit of the form $f_0 = \frac{df}{dP} \cdot P_{sg} + c$, where c is a constant, was computed. For comparison a power law fit of the form $f_0 = a \cdot P_{sg}^{\ b}$, with constants a and b, was also computed.

The means and standard deviations of the parameters a, b, $\frac{df}{dP}$ and c are given in Table 5-3 and Table 5-4, which shows that both fits are comparable in terms of R^2 error. Due to the scatter of the data it is not possible to infer any direct relationship between the level of applied tension and the fitting parameters. Hence, combined fits for all data of each sex were computed and the results included at the bottom of the respective table. The mean values over all data for both sexes were $a = 36 \pm 87\%$, $b = 0.48 \pm 61\%$, $\frac{df}{dP} = 0.75 \pm 104\%$ and $c = 190 \pm 149\%$.

In the literature, the correlation between P_{sg} and f_0 has been quantified with a linear approximation in most cases even though it is known to deviate from linearity in different ways depending on the subject (Plant and Younger, 2000). A linear fit of the form $f_0 = \frac{df}{dP} \cdot P_{sg} + c$, has two free parameters, however only $\frac{df}{dP}$ is of interest, while the offset c is generally not reported. However, in Table 5-3 and Table 5-4 c is shown to vary significantly.

In the data presented here, the one standard deviation range of $\frac{df}{dP}$ was 0.1-1.8 Hz/kPa for men's and 0.2-1.1 Hz/kPa for women's larynges, which are of the same order of magnitude as the typical literature values of 0.1-0.7 Hz/kPa (Alipour *et al.*, 2013; Alipour and Scherer, 2007; Titze, 1989). However, a single global linear fit could not take into account the very rapid changes in f_0 for small changes in P_{sg}, which have been observed with mean values of df/dP as high as 2.1 Hz/kPa here, or 2.0 Hz/kPa elsewhere (Alipour and Scherer, 2007; Lieberman *et al.*, 1969).

Table 5-3 Combined ranges of measurements and fitting data from all periods of phonation for female larynges under three levels of applied vocal fold tension for the air supply control parameter, and constant pressure under the mechanical control parameter.

		Air Supply Control			Mechanical Control
		High Adduction	Medium Adduction	Low Adduction	
LF1	P_{sg} (kPa)	0.3 - 9.9	0.8 - 8.3	0.3 - 7.7	1.0 - 7.5
	f_0 (Hz)	214 - 603	243 - 537	212 - 664	255 - 576
	a	11 ± 80%	28 ± 35%	41 ± 35%	21 ± 32%
	b	0.74 ± 91%	0.46 ± 21%	0.41 ± 24%	0.53 ± 19%
	R^2	0.95 ± 5%	0.98 ± 1%	0.96 ± 4%	0.98 ± 2%
	df/dP	0.65 ± 92%	0.50 ± 22%	0.60 ± 37%	0.75 ± 65%
	c	120 ± 203%	210 ± 21%	240 ± 25%	200 ± 26%
	R^2	0.95 ± 5%	0.97 ± 2%	0.96 ± 4%	0.98 ± 2%
LF2	P_{sg} (kPa)	0.1 - 4.6	0.1 - 5.6	1.0 - 3.5	1.2 - 2.3
	f_0 (Hz)	64 - 860	64 - 931	295 - 429	360 - 446
	a	5.2 ± 126%	6.0 ± 23%	68 ± 32%	61 ± 46%
	b	0.88 ± 28%	0.81 ± 8%	0.31 ± 34%	0.38 ± 44%
	R^2	0.95 ± 5%	0.99 ± 1%	0.89 ± 12%	0.92 ± 6%
	df/dP	1.7 ± 9%	1.6 ± 10%	0.52 ± 20%	0.82 ± 42%
	c	47 ± 84%	110 ± 39%	230 ± 12%	250 ± 23%
	R^2	0.94 ± 5%	0.99 ± 1%	0.91 ± 10%	0.92 ± 6%
LF3	P_{sg} (kPa)	1.8 - 9.7	2.9 - 8.1		0.1 - 9.6
	f_0 (Hz)	211 - 720	267 - 585		227 - 671
	a	3.5 ± 49%	3.1 ± 31%		14 ± 97%
	b	0.78 ± 17%	0.79 ± 12%		0.62 ± 45%
	R^2	0.99 ± 1%	0.99 ± 1%		0.85 ± 20%
	df/dP	0.59 ± 16%	0.61 ± 12%		0.51 ± 40%
	c	95 ± 61%	90 ± 45%		170 ± 70%
	R^2	0.98 ± 1%	0.99 ± 1%		0.85 ± 18%
LF4	P_{sg} (kPa)	1.4 - 2.7	1.2 - 3.0	0.4 - 2.2	1.5 - 3.3
	f_0 (Hz)	516 - 545	444 - 504	455 - 489	445 - 711
	a	310 ± 17%	220 ± 16%	410 ± 2%	120 ± 28%
	b	0.10 ± 53%	0.15 ± 32%	0.028 ± 21%	0.28 ± 34%
	R^2	0.79 ± 30%	0.95 ± 4%	0.94 ± 5%	0.95 ± 4%
	df/dP	0.26 ± 51%	0.35 ± 26%	0.13 ± 31%	0.67 ± 41%
	c	480 ± 6%	410 ± 6%	450 ± 2%	370 ± 14%
	R^2	0.79 ± 30%	0.95 ± 5%	0.91 ± 5%	0.95 ± 4%
All	P_{sg} (kPa)	0.1 - 9.9			
	f_0 (Hz)	65 - 931			
	a	50 ± 86%			
	b	0.45 ± 65%			
	R^2	0.95 ± 8%			
	df/dP	0.64 ± 65%			
	c	240 ± 56%			
	R^2	0.95 ± 8%			

Table 5-4 Combined data from all periods of phonation for male larynges, definitions are the same as in Table 5-3.

		Air Supply Control			Mechanical Control
		High Adduction	Medium Adduction	Low Adduction	
LM1	P_{sg} (kPa)	0.7 - 4.2		0.2 - 0.9	0.6 - 2.7
	f_0 (Hz)	240 - 520		116 - 300	298 - 516
	a	30 ± 99%		55 ± 41%	74 ± 27%
	b	0.64 ± 108%		0.4 ± 84%	0.32 ± 26%
	R^2	0.63 ± 61%		0.81 ± 6%	0.88 ± 11%
	df/dP	-1.4 ± 406%		2.1 ± 108%	0.68 ± 19%
	c	1100 ± 216 %		140 ± 60%	260 ± 155
	R^2	0.66 ± 47%		0.79 ± 11%	0.87 ± 12%
LM2	P_{sg} (kPa)	0.1 - 3.3	0.1 - 4.9	0.1 - 3.4	0.8 - 2.8
	f_0 (Hz)	102 - 329	151 - 356	131 - 313	122 - 338
	a	15 ± 73%	32 ± 58%	37 ± 47%	5.2 ± 35%
	b	0.57 ± 50%	0.41 ± 44%	0.4 ± 51%	0.75 ± 17%
	R^2	0.94 ±9%	0.93 ± 6%	0.89 ± 16%	0.94 ± 3%
	df/dP	0.79 ± 33%	0.46 ± 24%	0.70 ± 25%	1.0 ± 18%
	c	83 ± 65%	140 ± 30%	130 ± 33%	52 ± 53%
	R^2	0.94 ± 10%	0.92 ± 6%	0.91 ± 12%	0.93 ± 4%
LM3	P_{sg} (kPa)	0.1 - 2.9	0.1 - 2.4	0.1 - 1.7	0.6 - 3.9
	f_0 (Hz)	119 - 444	116 - 397	78 - 321	270 - 696
	a	66 ± 30%	49 ± 14%	51 ± 19%	39 ± 10%
	b	0.29 ± 39%	0.35 ± 19%	0.34 ± 23%	0.47 ± 7%
	R^2	0.89 ± 13%	0.96 ± 3%	0.96 ± 4%	0.99 ± 1%
	df/dP	0.65 ± 23%	0.86 ± 22%	0.99 ± 15%	1.2 ± 7%
	c	180 ± 14%	150 ± 11%	140 ±12%	230 ± 3%
	R^2	0.93 ± 7%	0.95 ± 3%	0.97 ± 2%	0.99 ± 1%
LM4	P_{sg} (kPa)	0.7 - 4.6	0.6 - 3.1	0.4 - 1.2	0.7 - 5.4
	f_0 (Hz)	154 - 490	166 - 344	101 - 210	150 - 538
	a	7.3 ± 42%	18 ± 42%	3.9 ± 88%	9.8 ± 60%
	b	0.73 ± 23%	0.54 ± 29%	1.0 ± 84%	0.69 ± 29%
	R^2	0.97 ± 3%	0.79 ± 25%	0.57 ± 39%	0.92 ± 8%
	df/dP	1.2 ± 43%	0.86 ± 32%	2.1 ± 80%	1.2 ± 50%
	c	96 ± 77%	110 ± 34%	2.9 ± 4394%	110 ± 68%
	R^2	0.96 ± 3%	0.79 ± 25%	0.57 ± 40%	0.92 ± 7%
All	P_{sg} (kPa)	0.1 – 5.4			
	f_0 (Hz)	78 – 696			
	a	26 ± 60%			
	b	0.50 ± 57%			
	R^2	0.91 ± 14%			
	df/dP	0.84 ± 117%			
	c	140 ± 247%			
	R^2	0.92 ± 13%			

The proposed power law of the form $f_0 = aP_{sg}^{\ b}$ in which both a and b are free parameters does not increase the complexity of the fitting and performs similarly to a linear fit $f_0 = \frac{df}{dP} \cdot P_{sg} + c$, when assessed in terms of the R^2 error, as shown in Table 5-3 and Table 5-4. Further, since the standard deviation of b is comparatively low (61% for all data combined) a simplification can be made to set $b = 0.5$, the mean value for all data, i.e. $f_0 = a\sqrt{P_{sg}}$. This empirical equation thereby reduces the number of free parameters for each larynx to one, a, which depends on the larynx. It is plausible that a is related to the applied tension, however, the scatter of the acquired data is too large to draw any conclusions on this point. The simplified relationship is shown as dashed lines in Figure 5-7 for various values of a. Such a fit to the data has the advantage of a level of prediction that is not present in the linear gradients typically used. This relationship suggests that the maximum rate of change of f_0 is dependent on the frequency, with more rapid changes possible at lower f_0. The lower mean values of $\frac{df}{dP}$ for women's compared to men's larynges may arise because the recorded phonation occurs at higher P_{sg}, where the rate of change is limited. Furthermore, for a purely dissipative model of vocal fold oscillation, in which all the kinetic energy of the air jet is lost, and therefore $\Delta P = P_{sg} = \frac{1}{2}\rho v^2$, f_0 would be seen to be proportional to v the air flow through the glottis.

5.2.4.Aryepiglottic larynx observations

For the preceding measurements, a Beckman-Eaton laminectomy retractor was used to separate the aryepiglottic larynx. A set of recordings made without the separation showed a passive narrowing of the aryepiglottic larynx, possibly due to the Bernoulli force, as the airflow through the glottis was increased. In some cases, vibration of the ventricular and/or aryepiglottic folds was observed. Figure 5-8 shows an example of the narrowing of the gap between the ventricular folds as P_{sg} was increased. These observations support the hypothesis that the approach and vibration of ventricular folds is governed only by aerodynamic effects (Bailly *et al.*, 2014, 2010, 2008).

The supraglottal laryngeal structures, in particular the ventricular folds, aryepiglottic folds and epiglottis can perform important linguistic roles (Esling *et al.*, 2005; Moisik *et al.*, 2010). However, control of the supra-laryngeal structures *in vivo* is not well understood and the extra thickness of the surrounding tissue may limit the narrowing of the aryepiglottic airway. Which muscles act to separate the ventricular and aryepiglottic folds and in which directions do they act?

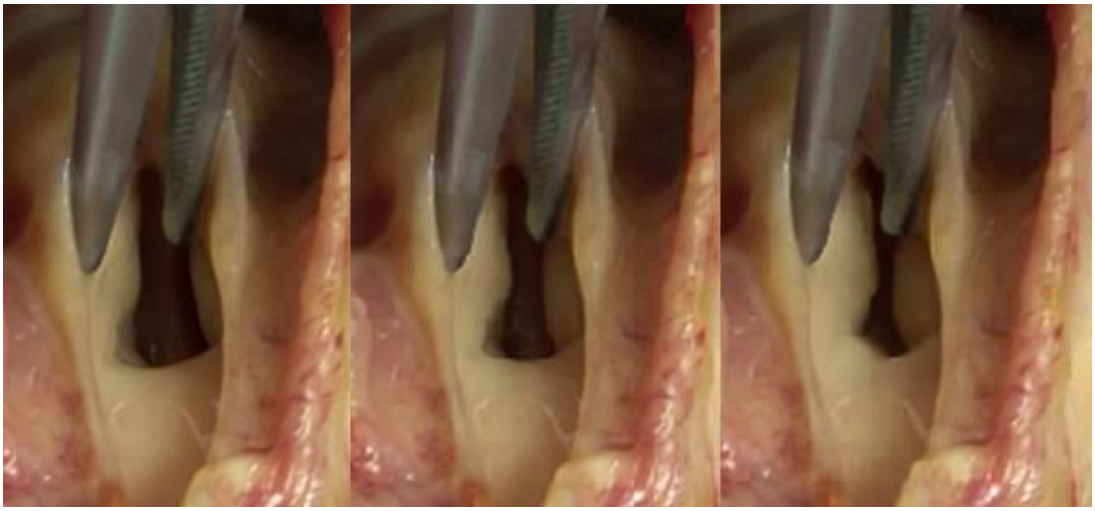


Figure 5-8 From left to right: three successive video frames showing the passive narrowing of the pale ventricular folds of LF1 as P_{sg} was increased. For clarity the aryepiglottic folds are held apart by the pliers visible in the upper part of the images.

Typically, supra-glottal structures are removed for the study of excised larynges (Alipour *et al.*, 2013; Mau *et al.*, 2011; McCulloch *et al.*, 2013). Here, the approach was to leave these structures intact but to separate them during measurements. Neither of these approaches simulates normal physiological conditions, so further work is needed to understand the properties of the aryepiglottic larynx. Although such work has begun on animal models (Alipour *et al.*, 2007), the animal anatomy is not directly comparable with that of humans.

5.3. Summary

Ex vivo control of the air supply and mechanical laryngeal parameters allows phonation over a wide range of f_0 and sound pressure level, even after freezing and thawing.

Reproducible discontinuities were observed either between harmonically related modes of vibration, or between modes of vibration with different glottal contact areas, analogous to laryngeal mechanisms M1 and M2. Estimations of the impedance load on the vocal folds showed that the discontinuities occurred in both inertive and compliant regions, suggesting that the behaviour is biomechanical and not triggered by acoustic loading.

The mean leap interval between mechanisms was recorded for one male larynx as 18 semitones for increasing glissandi and 15 semitones for decreasing glissandi.

Hysteresis was also observed as small pitch jumps with no change in glottal contact area. Hysteresis resembling visco-elastic behaviour was also observed in some cases.

For all larynges of both sexes, f_0 is approximately proportional to $\sqrt{P_{sg}}$, i.e. $f_0 = a \cdot P_{sg}^{0.5}$, where the empirical constant a is fitted for each larynx. Local linear gradients (df/dP) were found to be in the range 0.1-2.1 Hz/kPa comparable with literature values.

Airflow through the glottis narrows the aryepiglottic larynx.

6 *In Vitro*: Source-Filter Interaction

This chapter describes an investigation of partial occlusions of the vocal tract, similar to those used in speech therapy, performed on an *in vitro* vocal fold model[8]. Speech therapists sometimes ask patients to phonate through a small diameter straw at the (otherwise closed) lips. This therapy application is expected to have two different effects. First, it provides a significant DC resistance downstream from the glottis and therefore raises the pressure in the mouth, P_{io}, and therefore decreases the possible range of pressures across the glottis. Second, it modifies considerably the acoustic load acting on the glottal source. The Grenoble lab has a system of artificial vocal folds constructed by Ruty (2007) and the availability of this system meant that an experiment analogous to the speech therapy procedure could be conducted on an artificial system following the method of a preliminary study (Henrich and Ruty, 2012).

The moveable parts of the artificial 'vocal folds' are essentially pockets of latex filled with water. The mechanical properties of the folds are modified by adjusting the water pressure, making the folds more or less stiff.

6.1. Materials and Methods

6.1.1. *In vitro* setup

The experimental setup is a simplified replica of the human vocal tract, subglottal tract and vocal folds, designed to exhibit some physical phenomena relevant to human phonation in a situation with controllable parameters and simple geometry. The length of the subglottal tract was chosen to give a first resonance above the range of f_0 so that its influence on the vocal folds could be neglected and the acoustic effects of the vocal tract, which is comparable to that of an adult male, estimated. The latex vocal folds used are the "*Maquette 3*" described in (Ruty, 2007, p. 73). The design is said to be based on the work of Cronjaeger (1978), who suggested a triangular geometry in the plane of the vocal folds. Each vocal fold is constructed of a triangular metal structure with a central space. The structure is covered in a 0.2 mm thick latex sheet (Piercan, Paris, France) and the space is

[8]This work was performed at GIPSA-lab, University of Grenoble under supervision of Nathalie Henrich-Bernardoni, and with technical assistance from Xavier Laval. Preliminary results from this investigation were included in the invited presentation 'Phonation into straws: impact of an aeroacoustical load on in-vitro vocal fold vibration,' at the Pan European Voice Conference (PEVOC), Prague, Czech Republic (Hanna *et al.*, 2013b).

filled with water (see Figure 6-1). An outer layer of 0.5 mm latex is placed over the remaining uncovered parts of the structure and it is clamped in place to prevent leaks.

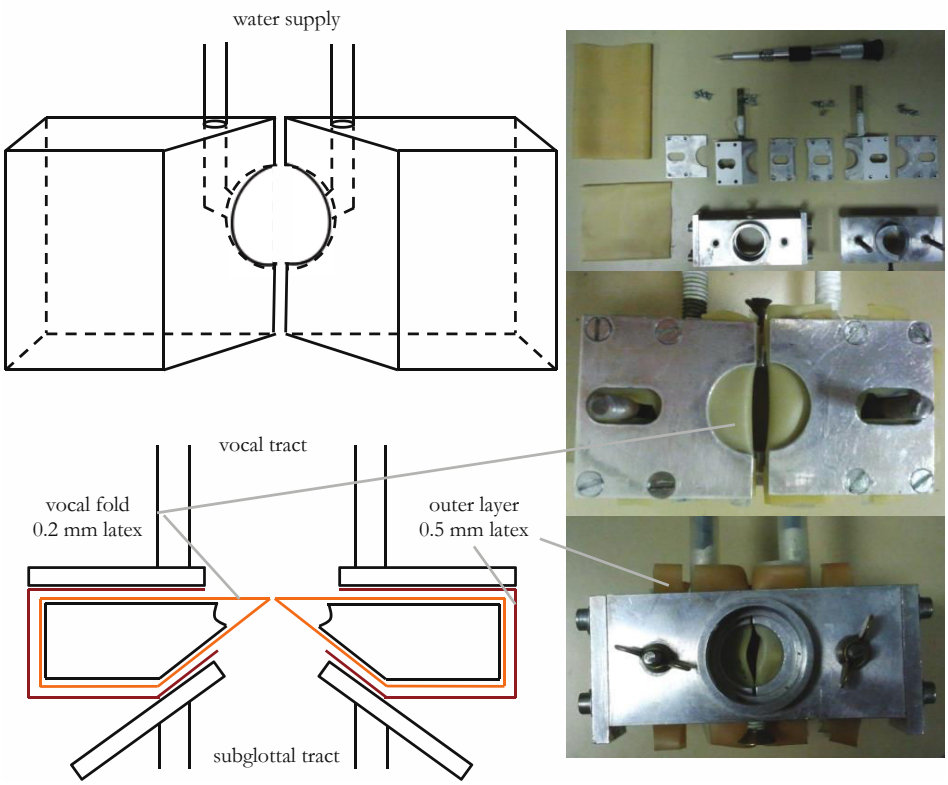


Figure 6-1 Schematic view of the structure that holds the latex vocal folds from upstream (top left) and above (bottom left). Photos showing the structure completely disassembled (top right), partially assembled with the 0.2 mm thick latex vocal folds visible (middle right) and completely assembled with a supporting layer of latex (bottom right) as viewed from downstream.

The rest of the setup is as described in (Ruty *et al.*, 2007). The apparatus consists of a 0.01 m^3 pressure reservoir filled with acoustical foam (*cf.* lungs) attached by a 9.5 cm subglottal tract to a latex vocal fold replica and 18 cm cylindrical supraglottal vocal tract model as shown in Figure 6-2. The vocal tract is either open at its end with no straw, a diameter of 25 mm, or partially occluded by a baffle which is traversed by 150 mm long straws of diameter 9, 6, or 4.5 mm respectively, as shown in Figure 6-2.

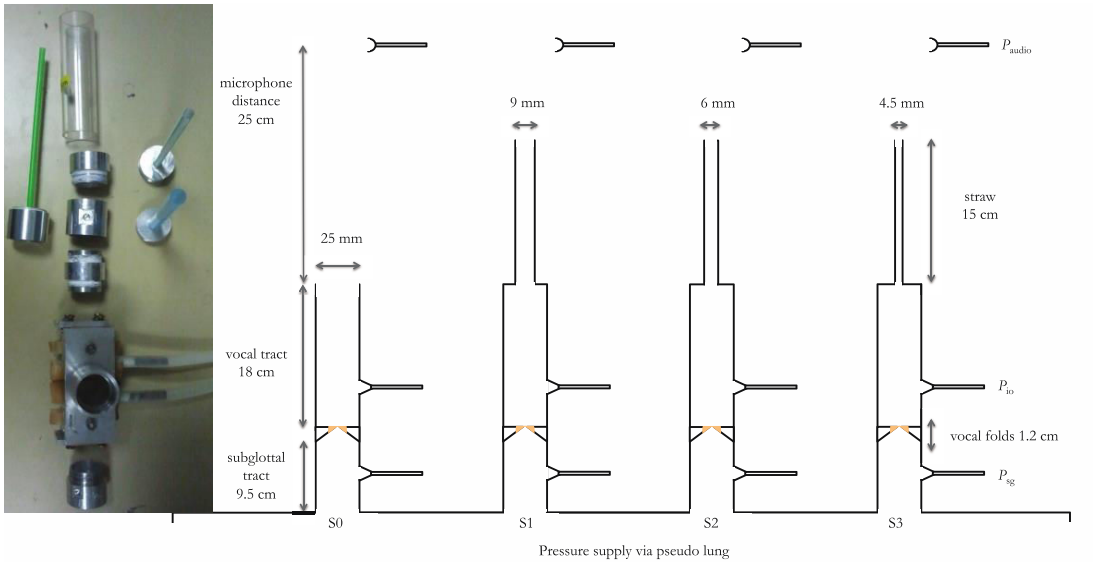


Figure 6-2 Schematic representation of the experimental setup for the four cases in each trial. Only the termination of the vocal tract is different in each of the four cases, where it is either open (S0) or an airtight connection is made to a straw of 15 cm length and diameter 9, 6 or 4.5 mm (S1-3).

Both the air pressure upstream of the replica (P_{sg}) and the water pressure within the replica (P_w) are adjustable. P_w was determined by a water reservoir above the axis of the vocal tract. Each 10 cm increase in height corresponds to an increase of ~980 Pa of water pressure inside the latex vocal fold replica. The air pressure upstream and downstream of the vocal fold replica was measured using pressure transducers (XCS-093, Kulite, Leonia, NJ, USA). The audio signal was recorded with a microphone and preamplifier (B&K 4192 and B&K 5935L, Brüel & Kjær, Nærum, Denmark). The opening of the vocal folds during excitation was monitored by an upstream laser focussed on a light sensitive diode downstream using the same equipment and method as Vilain *et al.* (2003). All signals were recorded using a PCI-MIO-16E4 digital acquisition card (National Instruments Corporation, Austin, TX, USA) and LabView software. Additionally, in one set of experiments (Trial III), the airflow at the downstream termination of the vocal tract was measured using an aerodynamic workstation (EVA2, S.Q. Lab, Aix-en-Provence), connected directly to the vocal tract for the open tract case, and via 15 cm tubes of comparable diameter for the straw cases.

The mechanical resonances of the vocal folds were measured in two sets of experiments (Trials II and III) with the vocal tract replica removed. Following the method of Hermant *et al.* (Hermant *et al.*, 2012; Silva *et al.*, 2014) a 30 second broadband excitation from 50-

400 Hz, made up of sine waves spaced at 1.22 Hz (Smith, 1995) was applied to the vocal folds via a shaker amplifier (TMS K2004E01) connected by a metal rod. An accelerometer (PCB 353B18) attached to the rod was used to record the frequency response of the system along with the time varying vocal fold aperture from the laser signal. Measurements were made with the excitation at four different points on the latex for each water pressure level to elicit potentially different mechanical modes of vibration available to the folds. The setup is shown in Figure 6-3.

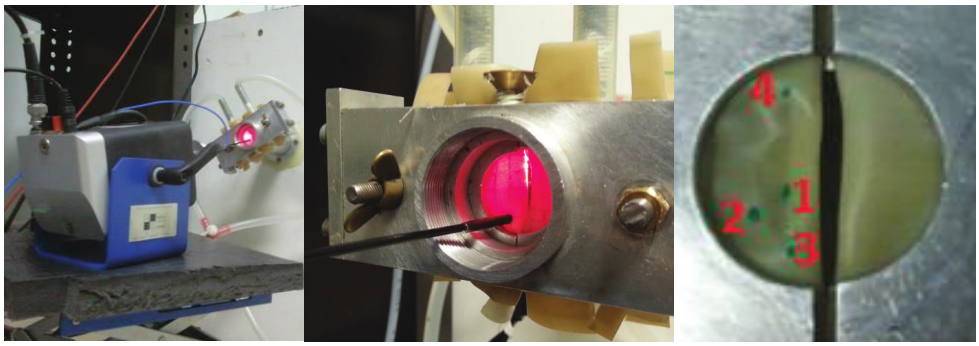


Figure 6-3 Photos showing the shaker amplifier applying a mechanical excitation to one of the four positions on the latex vocal folds. The four positions are shown on the right (with the folds deflated).

In Trial II the mechanical response of a single vocal fold was measured while the other was deflated. In Trial III the mechanical resonances were measured with the two vocal folds inflated, which yielded a different set of resonances including combinations of the two folds, which are not directly comparable to those in Trial III.

6.1.2.Experimental protocol

6.1.2.1. Trial I

A pilot experiment was performed by Henrich and Ruty (2012) and the resulting data are hereafter referred to as Trial I. The protocol for Trial I was as follows: the height of the water reservoir above the tract axis was fixed, then P_{sg} was increased by supplying pressurised air through the lung replica until the onset of audible oscillation of the latex vocal folds. P_{sg} was fixed at that level for several seconds to allow the oscillation to stabilise, before being gradually decreased. Three of these measurements were made with each vocal tract configuration, then the height of the reservoir was increased by 5 or 10 cm and the process repeated.

6.1.2.2. Trial II

For Trial II, new latex was installed into the vocal fold replica and the Trial I protocol was followed. After reaching the maximum height (after which oscillation was impossible), the process was resumed in reverse, decreasing the height of the reservoir by 10 cm each time. The latex replica demonstrates some hysteresis due to stretching, so there were differences in the results for the ascending and descending measurements. The mean of the values was computed.

6.1.2.3. Trial III

The protocol for Trial III followed that of Trial II but to avoid some of the hysteresis effect, the latex vocal fold model was one that had been used previously for an unrelated experiment, and so effectively 'broken in'. The mean of the ascending and descending data was computed. For the no straw case, the tract was connected directly to the EVA2 aerodynamic station to measure the airflow, this places an unknown acoustic load on the vocal folds but does not affect the DC pressure. For the straw cases, the attachment was made via 15 cm flexible tubes of appropriate diameter. This effectively lengthens the straws compared to Trials II and III.

6.1.3.Data analysis

For each of the three trials, three measurements were made at each P_w for each of the four conditions: no straw, and straws of 9, 6 and 4.5 mm diameter. The collected data were processed in Matlab. The onset and offset times were determined using the YIN f_0 estimation (de Cheveigné and Kawahara, 2002) and these times used to identify the corresponding subglottal, intra-oral, and audio pressures P_{sg}, P_{io} and P_{audio}, respectively. The onset pressure required to start oscillation is typically higher than the offset pressure and f_0 changes rapidly close to the onset and offset. To remove these transient parts of the phonation, the median f_0 value between onset and offset was selected to ensure that the stable part of the oscillation was chosen. The raw data collected of Henrich and Ruty (2012) were reanalysed in the same way so that they were directly comparable.

6.1.4.Acoustic impedance calculations

The acoustic impedance of the system as seen by the vocal folds was calculated using the approach described in Chapter 5. The load on the vocal folds is that of the vocal tract, which in turn is loaded by either the radiation impedance, or that of open pipes of length 150 mm and diameters of 9, 6, and 4.5 mm. The vocal tract is considered as rigid, to

correspond to the *in vitro* experiment; and with non-rigid walls, to correspond to the *in vivo* application of straw therapy.

6.2. Results and Discussion

6.2.1. The P_w, f_0 relationship

The internal water pressure P_w of the vocal folds acts somewhat like the mechanical tension adjustments used in chapter 5. Increasing P_w increases f_0, and at a given P_w, f_0 decreases with decreasing straw diameter, as shown in Figure 6-4.

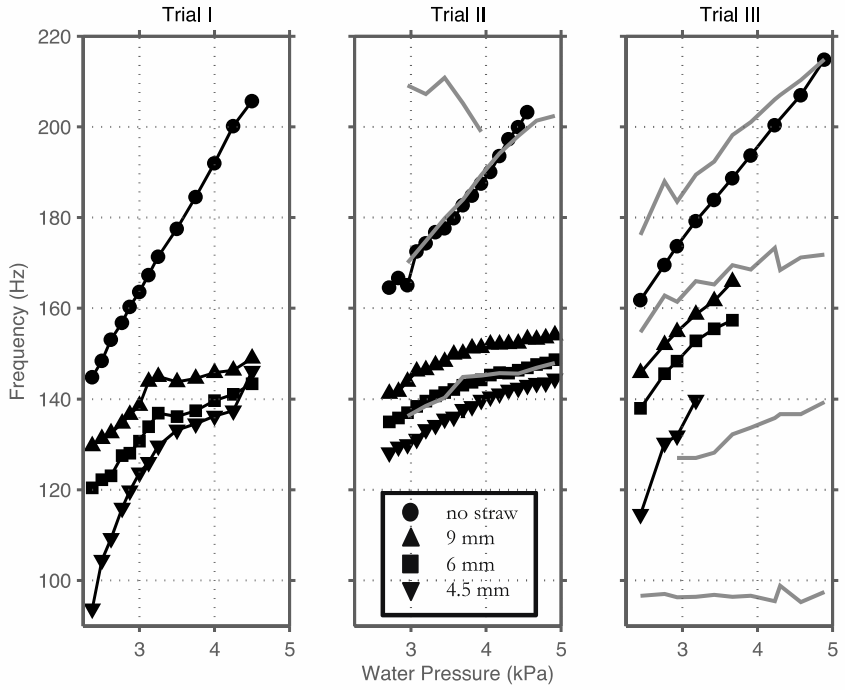


Figure 6-4. Dependance of f_0 on P_w in the three trials. Mechanical resonances are shown as pale solid lines.

A disadvantage of this system is that the stiffness of the latex is determined by the relative pressure between the vocal folds and the air in the vocal tract, i.e. P_{io}. When straws are used $P_{io} \neq 0$, so that the actual relative water pressure is not accurately known.

6.2.2. Mechanical resonances of the vocal folds

The excitation from the shaker amplifier was applied to four separate locations on the latex vocal folds (shown in Figure 6-3). If the location excited corresponds with a vibration node, the associated resonance may not be observed. Figure 6-5 shows an example of the four

response spectra at the four locations made with the same internal water pressure. The principal modes of vibration were determined from the mean of the mechanical response maxima and zero crossing frequencies of the phase over the four excitation positions. Less prominent vibrations that are observed in only one position therefore contribute little to the measurements.

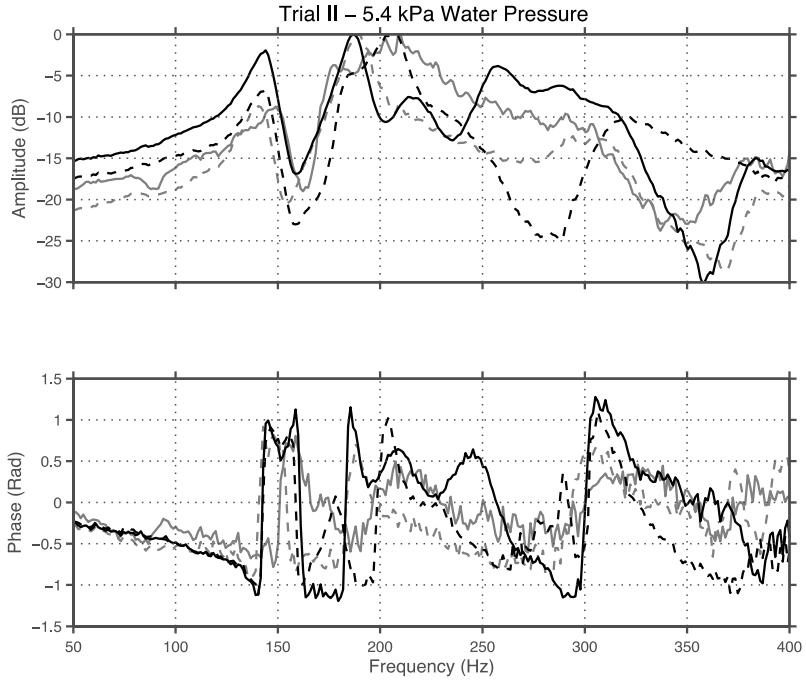


Figure 6-5. Example broadband excitation response showing normalised glottal aperture from the photodiode laser signal with excitation at the four different positions with the same vocal fold internal water pressure. In the example shown, the first resonance at 140 Hz is excited by all four of the positions. *cf.* mechanical excitation *in vivo* in Figure 2-5 reproduced from Švec *et al.* (2000).

Typically, there were three prominent resonances in the measured 50-400 Hz range, which increase with P_w, as shown in Figure 6-4. These are in qualitative agreement with the three peaks at approximately 110, 170, and 240 Hz for f_0 = 110 Hz reported *in vivo* (Švec *et al.*, 2000) (see Figure 2-5).

In Trial II, Figure 6-4 shows that for the no straw case, f_0 coincides with the highest mechanical resonance of the latex vocal folds. For the straw cases, f_0 lowers by 20-60 Hz to approach the next highest mechanical resonance. No mechanical resonance measurements were made during Trial I, and the measurements in Trial III yield slightly different behaviour due to the excitation of cross-modes between the two vocal folds.

6.2.3.Vibratory characteristics in an open vocal tract configuration

Since $P_{io} = 0$, the open vocal tract configuration is somewhat comparable to the *ex vivo* experiments in chapter 5. To first order, the P_{sg}, f_0 relationship, shown in Figure 6-6, could be considered linear but Table 6-1 shows that the R^2 values for linear fits are consistently lower than power law fits.

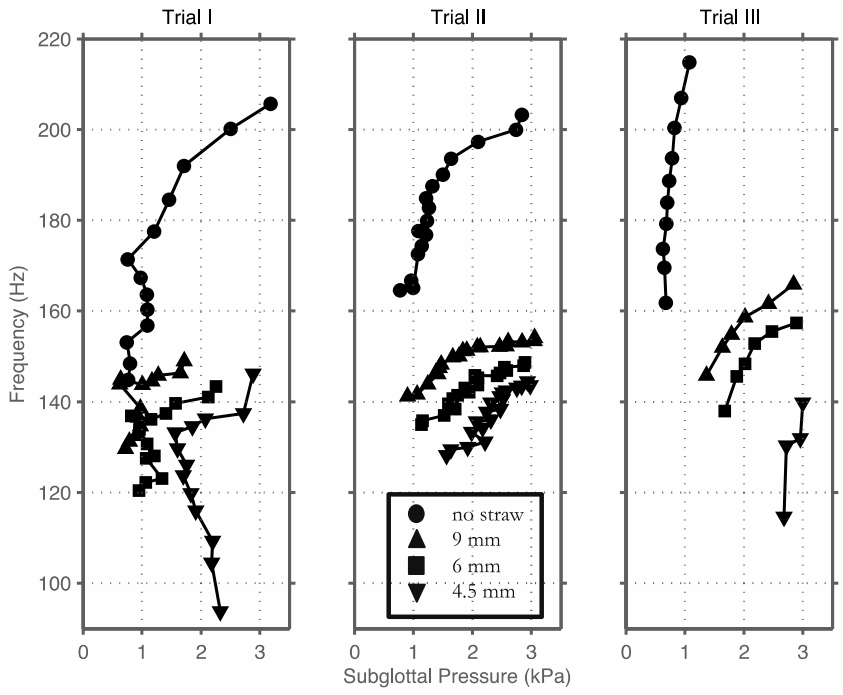


Figure 6-6 The dependence of f_0 on P_{sg} in the three trials.

Although the latex replica is designed to illustrate qualitatively the self-oscillation of the vocal folds, the mechanical properties of an elastic tube filled with water are not sufficiently similar to those of human tissues to expect that the P_{sg}, f_0 relationship would be quantitatively similar. Nevertheless, in Trial III, f_0 is well fitted by a function proportional to $P_{sg}^{0.5}$, which is similar to the mean *ex vivo* experiments discussed in Chapter 5.

Table 6-1. Modelling parameters for the P_{sg}, f_0 relationship with an unobstructed vocal tract (*cf. ex vivo* data in Table 5-3 and Table 5-4)

Trial		Increasing	Decreasing	Mean
I	P_{sg} (kPa)	0.7 - 3.2		
	f_0 (Hz)	145 - 206		
	a	160		
	b	0.22		
	R^2	0.82		
	df/dP	23		
	c	140		
	R^2	0.78		
II	P_{sg} (kPa)	1.0 - 2.8	0.8 - 2.8	0.8 - 2.8
	f_0 (Hz)	160 - 203	165 - 200	165 - 203
	a	180	170	170
	b	0.15	0.17	0.17
	R^2	0.70	0.95	0.87
	df/dP	16	19	18
	c	160	150	160
	R^2	0.66	0.88	0.79
III	P_{sg} (kPa)	0.6 - 0.9	0.7 - 1.2	0.6 -1.1
	f_0 (Hz)	161 - 213	162 - 216	162 - 215
	a	230	200	210
	b	0.60	0.38	0.47
	R^2	0.75	0.87	0.86
	df/dP	150	81	110
	c	77	120	100
	R^2	0.75	0.85	0.84

Table 6-1 shows that the P_{sg}, f_0 relationship in Trials I and II are very similar, with f_0 proportional to $P_{sg}^{0.2}$ in both cases. Trial III requires a larger constant *b* or linear gradient *df/dP*.

The oscillation behaviour of the latex vocal folds is influenced by the mechanical resonances of the latex (section 6.2.2) and the downstream impedance (section 6.2.6). The impedance would be inertive in the region of f_0 for both the *in vitro* and *ex vivo* experiments, although a larger impedance is expected in this experiment due to the vocal tract replica in contrast to the radiative impedance of the open aryepiglottic larynx in the *ex vivo* experiments. Here, the variation between Trials I, II, and III is likely due to differences in the initial latex parameters, e.g. thickness, which vary within the manufacturers specification, and their properties, e.g. stiffness, which depend on the mounting of the replica and vary with age and use.

6.2.4.Influence of transglottal pressure on f_0

For the no straw case, the transglottal DC pressure $\Delta P = P_{sg}$-$P_{io} = P_{sg}$, since $P_{io} = 0$, hence the discussion of the P_{sg}, f_0 relationship above. For the straw cases, $P_{io} \neq 0$ as shown in Figure 6-7, so that the ΔP, f_0 relationship is more likely to drive the vocal fold oscillation.

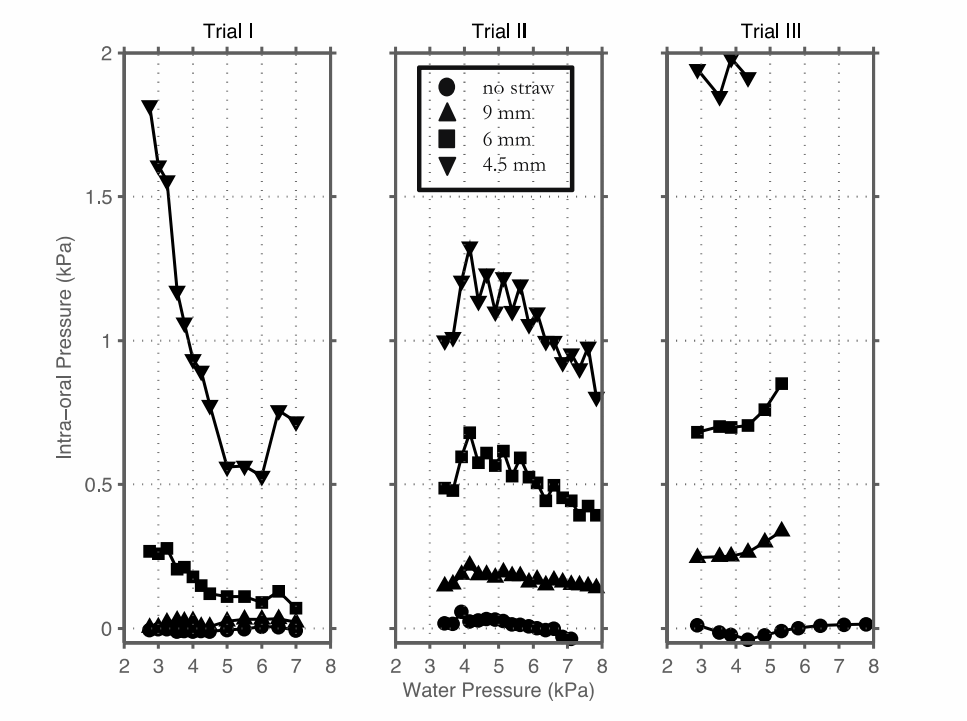


Figure 6-7. The dependence of f_0 on P_{io} in the three trials.

Figure 6-8 shows the ΔP, f_0 relationship for the three trials. All trials show that f_0 increases with ΔP. Furthermore, for the straw cases, f_0 appears almost independent of the straw diameter. Note, that the relationship for the no straw case is almost identical to that shown in Figure 6-6 since $P_{io} = 0$ kPa.

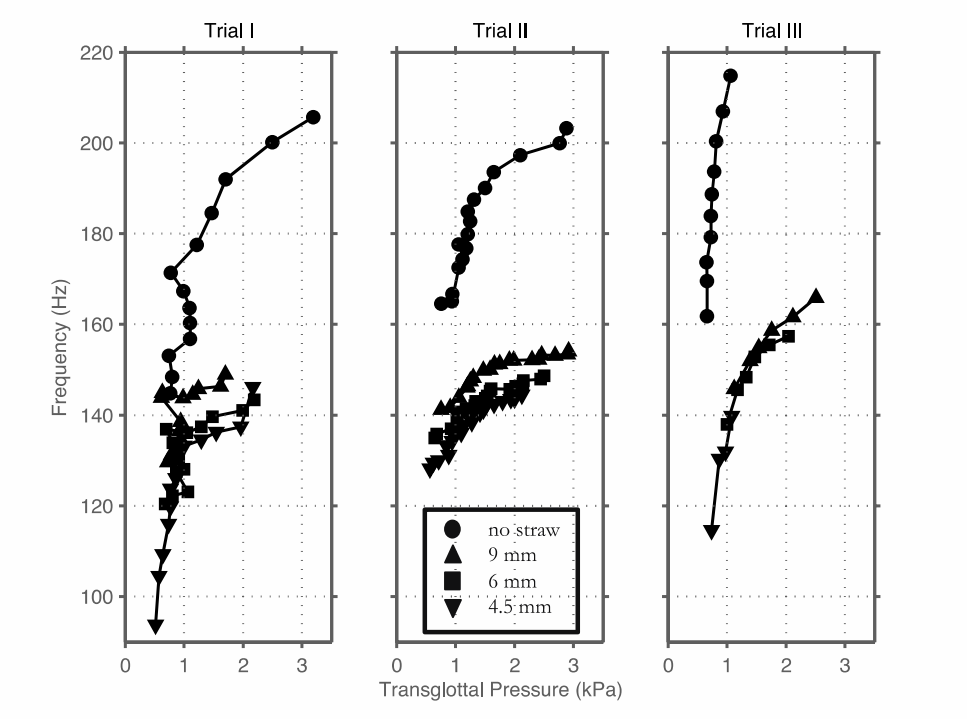


Figure 6-8. The dependence of f_0 on ΔP in the three trials.

6.2.5. Aerodynamic effects

In Trial III the airflow Φ at the output of the vocal tract or straw was measured. Increasing P_w affected Φ in a different way depending on the downstream load. Φ decreases with increasing P_w for the no straw case, but Φ increases with increasing P_w when a straw is in place as shown in Figure 6-9. High speed video recordings appear to show that for the no straw case the maximum glottal area decreases with increasing P_w, as shown in Figure 6-10. However, without quantifying the area as a function of the period of oscillation, it is not clear whether this is related to the decrease in Φ.

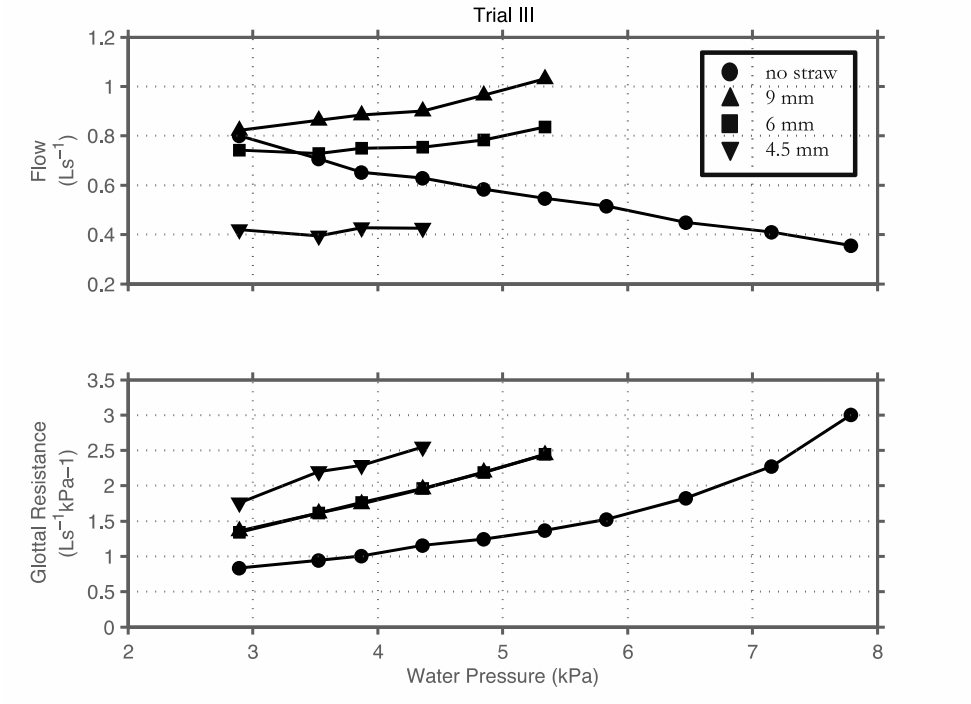


Figure 6-9. Airflow Φ as a function of P_w, and glottal resistance as a function of P_w.

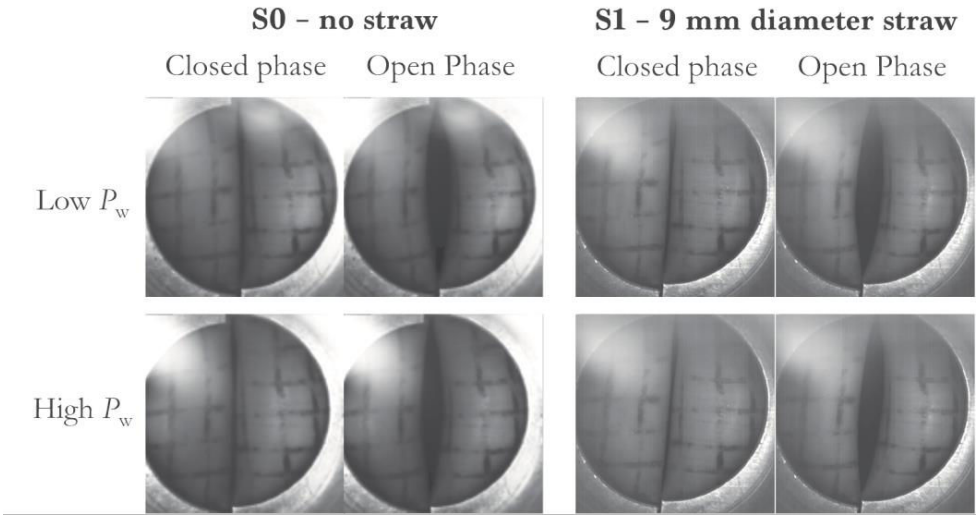


Figure 6-10 Still frames extracted from high speed video recording show the minimum and maximum glottal area during the closed and open phases of the oscillation. In the no straw case, there appears to be a decrease in glottal area with increasing P_w, which may be related to the decrease in airflow as a function of P_w shown in Figure 6-9.

The measured airflow allows consideration of the glottal resistance $\Delta P/\Phi$, also shown in Figure 6-9. The effect of the partial constrictions is to raise the glottal resistance. The smallest (4.5 mm) diameter straw has the highest glottal resistance.

The apparent difference in behaviour of the no straw case could be due to the much steeper $\Delta P/P_w$ gradients for the straws than for no straw in Trial III when the airflow was measured, as shown in Figure 6-11. This difference in gradients is not so large in Trials I and II. Since the aerodynamic station did not increase P_{io} for the no straw case (see Figure 6-7), therefore the DC load is unchanged, perhaps the difference in behaviour is caused by the unknown acoustic load impedance of the measurement device.

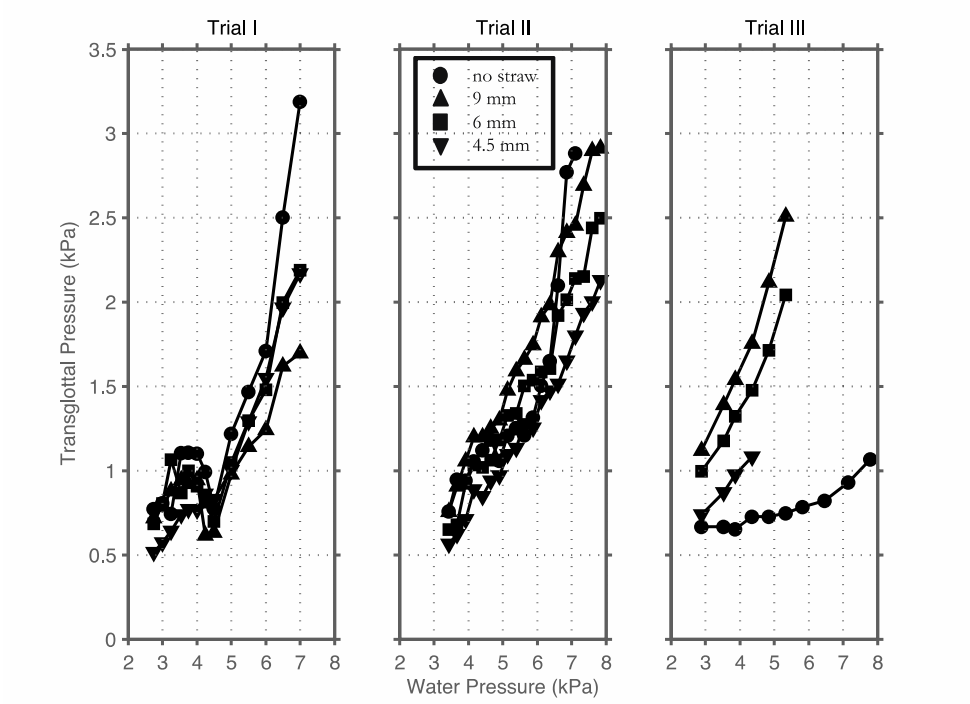


Figure 6-11. The dependence of ΔP on P_w in the three trials.

The main influence of the vocal tract occlusion with the straw on the vocal folds is aerodynamic. The data here are in agreement with the *in vitro* and numerical models of Bailly *et al.* (2008) who demonstrated that a downstream constriction of the vocal tract affects vocal fold vibration by changing P_{io} and therefore ΔP. In addition, the acoustic properties of the vocal tract may also have some influence.

6.2.6. Acoustic effects

6.2.6.1. Modelling the impedance on the in vitro vocal fold replica

In addition to aerodynamic factors, some effect on vocal fold motion under phonation into a straw may be due to the acoustic load 'seen' by the vocal folds. Figure 6-12 shows a simple rigid waveguide model using the dimensions of the *in vitro* setup, following the same method as in section 4.2.1.4. Note that the short subglottal tract of length 9.5 cm was chosen so that its first resonance would be above the range of f_0 (90-220 Hz). Therefore the vocal tract and constriction are considered as the load on the vocal fold and the inertance of the subglottal tract is not considered.

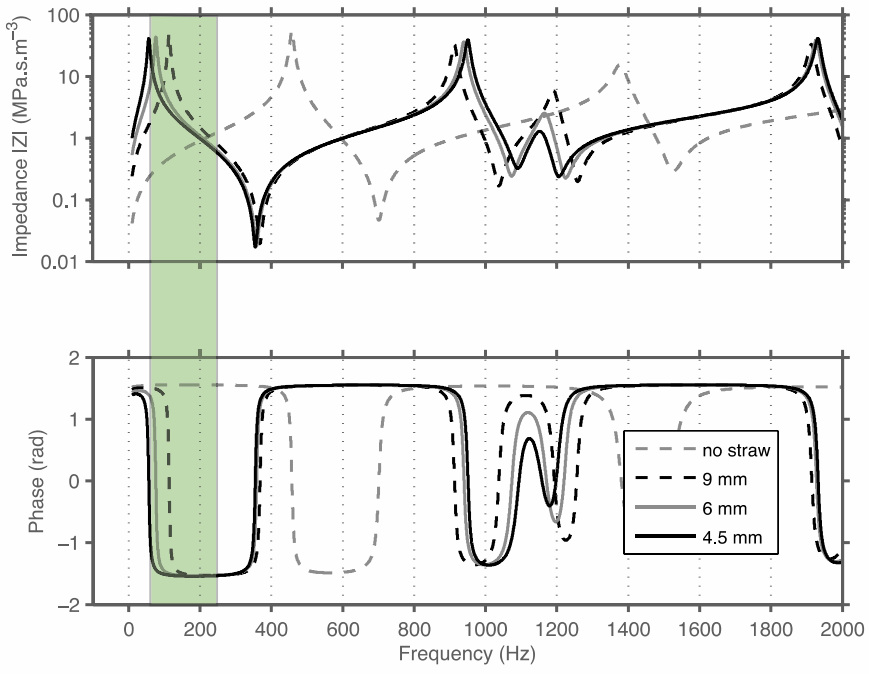


Figure 6-12. (top) Acoustic impedance magnitude for the *in vitro* setup as 'seen' from an effective glottal radius of 5 mm, calculated using the method described in section 4.2.1.4. The range of f_0 achieved in this experiment is highlighted. The scale is extended for comparison with the closed lip configuration in Figure 4-5). (bottom) Phase of the impedance. A non-rigid model is shown in Figure 6-13.

The effect of such an occlusion is to add a Helmholtz resonance at a frequency between 50 and 110 Hz, i.e. lower than the typical frequency of phonation. Therefore the vocal folds 'see' a compliant acoustic impedance when a straw is used, but an inertive impedance when the vocal tract is unobstructed. Such a difference in the acoustic load on the vocal folds may influence the motion of the vocal folds, and perhaps encourage the change to the

lowest mechanical mode of the folds. However, *in vivo* the non-rigid walls of the vocal tract must be considered.

6.2.6.2. Modelling the impedance on the vocal folds

As shown in section 4.2.1.4.3, the inertance of the non-rigid walls of the real vocal tract place a lower limit on the lowest formant of the voice of ~200 Hz. At low frequency, the inertance of the walls and the inertance at the lips are in parallel, so there is a limit to the effect of increasing the inertance at the lips by lengthening the vocal tract or constricting it. So it is not possible to reduce the first resonance below the mechanical resonance of the walls. This mechanical wall resonance is not present in the rigid *in vitro* model, so the constriction of the vocal tract by a straw can introduce a low frequency resonance below f_0, turning the load seen by the vocal folds from inertive to compliant. However, with non-rigid walls the effect of increasing the inertance at the lips will force the first impedance maximum as seen by the vocal folds to be no lower than ~200 Hz, therefore within the range of f_0, rather than below it as shown in Figure 6-13.

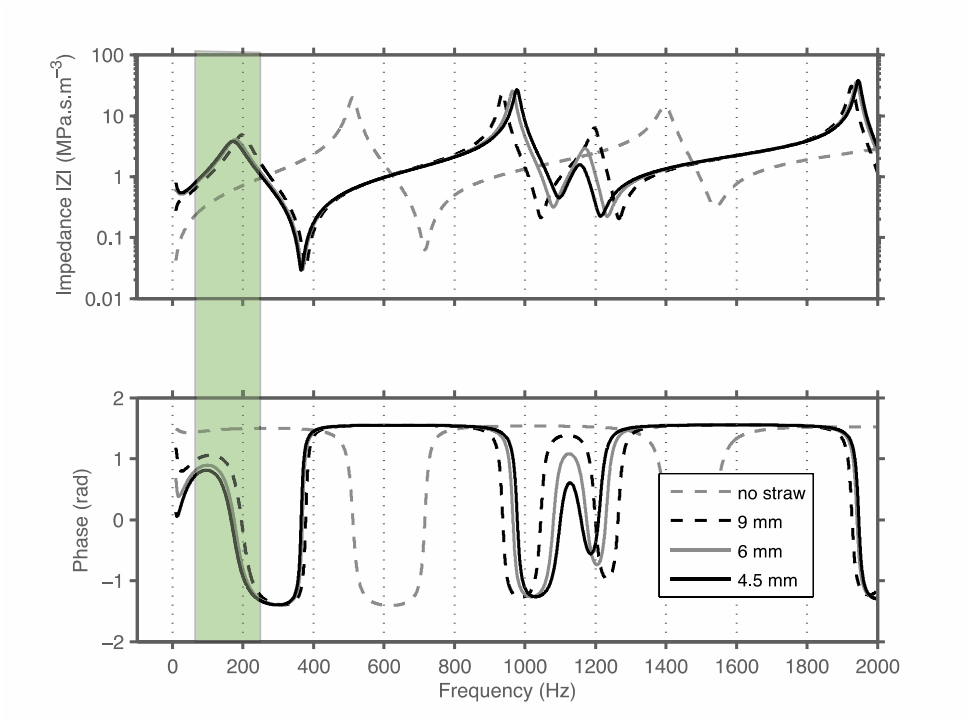


Figure 6-13 Impedance magnitude and phase as in Figure 6-12, except that the non-rigid vocal tract parameters from chapter 3 were used for the vocal tract (*cf.* the closed lip configuration in Figure 4-5). Note that the extra inertance of the smallest (4.5 mm) straw does not lower the frequency of the first maximum below that of the 6 mm diameter straw, i.e. the closed vocal tract limit is achieved, *cf.* lowest maxima in Figure 6-12.

It is therefore possible that one of the effects of straw phonation is that the vocal folds experience a larger than normal inertance, which encourages the swinging door motion of the folds during phonation (Titze, 2008). However, for high f_0 the load may appear compliant which may be less favourable, therefore care must be taken in selecting a straw with appropriate inertance for the subject's f_0.

6.3. Summary

The continuing relevance of the *in vitro* water-filled latex vocal fold replica of Ruty (2007) has been demonstrated, with a P_{sg}, f_0 relationship and mechanical resonances comparable with those of the human vocal folds measured *ex vivo* (in Chapter 5) and *in vivo* (Švec *et al.*, 2000) respectively.

When the vocal tract is partially occluded by a straw the replica demonstrates evidence of source-filter interaction; f_0 is influenced by the aerodynamic pressure, the mechanical resonances of the vocal fold model, and the downstream acoustic load.

In summary, a partial occlusion of the vocal tract by the straws in these experiments:

- increases the intra-oral DC pressure and decreases the subglottal pressure for the same water pressure.
- causes the airflow to increase with increasing P_w, whereas the airflow decreases with increasing P_w with an open 'mouth'.
- increases the glottal resistance, i.e. the ratio of transglottal pressure:flow.
- decreases f_0 by 20-60 Hz, from coinciding with a high mechanical resonance of the vocal fold replica, to a lower one.
- changes the downstream acoustic load from inertive to compliant in the *in vitro* model in the region of f_0. This effect, however, may have limited relevance to a real vocal tract in which the non-rigid walls place a lower bound on the impedance maximum, although it may still be the case for high f_0. Otherwise, the constriction increases the inertance.

7 Conclusions

7.1. Summary of Experimental Chapters

Chapters 3 and 4 reported measurements of the acoustic impedance spectra of the vocal tract made from 10 Hz to 4.2 kHz by injecting a synthesised broadband acoustic signal between the lips. With the glottis closed, the acoustic resonances of the closed tract between about 300 Hz and 4 kHz, which are responsible for the first three voice formants, have impedance magnitudes of 0.1 MPa·s·m^{-3} and anti-resonances (impedance peaks) have impedances of 10 MPa·s·m^{-3}. In the absence of radiation at the open lips, the bandwidths of the acoustic resonances increase with increasing frequency from ~50 to 90 Hz for men and ~70 to 90 Hz for women for both resonances and anti-resonances. These bandwidths are consistent with losses five times greater than the visco-thermal losses of a smooth rigid cylinder of comparable dimensions. The pressure transfer function between glottis and lips was estimated for a similarly simple model and the values obtained are in agreement with formant bandwidth estimates in the literature. An open glottis has a finite acoustic impedance, as opposed to the infinitely large impedance for the closed glottis. During the measured phonation, the air in the glottis acts as a lossy, inertive impedance, increasing the bandwidths of the resonances but barely changing the resonant frequencies.

Acoustic and electromagnetic measurements reveal a resonance and anti-resonance below 300 Hz, and the measured data give an estimate of wall inertance that is consistent with a (hypothetically uniform) wall thickness of 1-2 cm. The first impedance minimum falls at about 20 Hz and comprises the tract walls vibrating on a 'spring' of distributed stiffness per unit area of 2-4 kN·m^{-3}. The first impedance maximum, as measured from the lips, is at about 200 Hz and increases in frequency with increasing glottal aperture. This maximum determines the lower limit of the first formant, which is approached by reducing the effective lip aperture nearly to zero.

For an effective glottal radius > 1 mm, the impedance of the subglottal vocal tract affects the impedance measured at the lips. This impedance is reasonably well modelled by the addition of a cylinder of length ~20 cm open at the far (lung) end, that provides an acoustic load at the glottis. For breathy phonation, the resonances of the subglottal tract may influence vocal fold motion depending on whether their response is modelled as a series or parallel load on the glottal source.

Chapter 5 described experiments on several excised human larynges, in which no supraglottal vocal tract was present. The subglottal plumbing present to supply a source of pressurised air does provide an acoustic load that may influence the vocal fold motion. Vocal fold oscillation was achieved over a wide range of P_{sg} and vocal fold tensions. In general, increasing either the aerodynamic or mechanical control parameters caused an upward glissando (pitch glide), and decreasing either caused a downward glissando. Four types of discontinuous or hysteresis behaviour were encountered: intra- and inter-mechanism pitch jumps, discontinuous changes to subharmonic oscillations, and visco-elastic hysteresis. Disregarding such discontinuities, f_0 was found to be approximately proportional to $\sqrt{P_{sg}}$. Further, the fluid-structure interactions within the larynx were found to induce a narrowing of the aryepiglottic larynx, and in some cases trigger vibration of the ventricular and aryepiglottic folds. This highlights the lack of knowledge of the *in vivo* control of the aryepiglottic larynx and how it is simulated *ex vivo*. No clear evidence of the subglottal plumbing influencing the behaviour of the vocal folds was observed.

In chapter 6, a series of straws were placed at the 'lips' of an *in vitro* water-filled latex vocal fold replica. This simulates a speech therapy exercise in which subjects phonate through straws placed between the lips. The technique modifies the acoustic and aerodynamic properties of the vocal tract as 'seen' by the vocal folds, and thus affects their vibration. In these experiments, the subglottal plumbing was short so that the f_0 always occurred on the inertive side of the resonance, and its impedance magnitude is small, so that it was expected to have little influence on the vocal fold motion.

Oscillation of the vocal fold replica was driven by an adjustable subglottal pressure, and its mechanical properties were determined by the pressure of water inside the latex 'skin' of the vocal fold replica. Increasing the water pressure causes the folds to stiffen and change the frequency of its mechanical resonances. The replica was connected to a rigid vocal tract model, terminated by an opening at the 'lips' to the air, or partially occluded with one of three straws.

For the open lip case, f_0 was comparable with the highest of the determined mechanical resonances of the vocal folds. Placing a straw at the termination decreased the transglottal pressure and caused f_0 to approach the next highest mechanical resonance. This demonstrates that the fundamental frequency of the vocal fold oscillation is affected not

only by the stiffness (internal water pressure) but also by the transglottal pressure, aerodynamic and acoustic loads and the mechanical resonances of the folds.

The straws also change the acoustic load on the vocal tract from inertive to compliant in the range of f_0, which although likely a secondary effect and not necessarily applicable to a non-rigid vocal tract, may also influence vocal fold vibration.

7.2. Conclusions

Although the acoustic resonances of the neutral vowel are reasonably approximated by those of a rigid cylinder, the attenuation coefficient of the vocal tract is roughly five times greater than the standard visco-thermal losses of a rigid cylinder of comparable dimensions.

The mass of the walls of the tract and the spring constant of its supporting tissues contribute a mechanical resonance and anti-resonance to the acoustic impedance measured through the lips, at about 20 and 200 Hz respectively.

The anti-resonance at ~200 Hz, due to the spring of air in the tract in parallel with the mass of the tract walls provides a lower limit to the first formant frequency. This limit is approached when the vocal tract is partially occluded, as is the case with closed vowels and when straws are placed between the lips. Therefore a non-rigid vocal tract model is necessary for accurate modelling of the impedance seen by the glottis and the vocal tract transfer function for these cases.

Measurements at the lips show that the impedance change between phonation and closed glottis is consistent with an effective average glottal radius < 1 mm. For high glottal impedance (small aperture), the impedance spectrum seen by the vocal folds has strongly asymmetric extrema; minima occur at frequencies only slightly above those of the maxima. Therefore the maxima in the pressure transfer function from the lips to the glottis occur close to the closed glottis anti-resonances measured through the lips.

Decreasing the glottal impedance makes the impedance spectrum more symmetric, so that the formant frequencies and bandwidths are increased, as has been reported for whispered speech.

For an effective glottal radius > 1 mm, the effects of the subglottal tract are observed in the impedance spectrum measured at the lips. The observed changes to the impedance

spectrum can be accounted for by a simple model of the subglottal tract as a ~20 cm cylinder open at the far (lung) end.

The importance of the subglottal impedance to vocal fold vibration is an open question. Both series and parallel combinations of the up- and downstream impedances have been suggested for the load on the glottal source, for different mechanisms of phonation. Models of both combinations show that the subglottal tract contributes additional peaks to the impedance spectrum.

Using only a subglottal tract and no vocal tract, a wide range of fundamental frequencies and sound pressures can be achieved with excised larynges by adjusting the supplied subglottal pressure, or by fixing the air supply and adjusting the mechanical properties of the vocal folds. In both cases, the fundamental frequency was found to be proportional to the square root of the subglottal pressure.

Four types of discontinuous or hysteresis behaviour were observed in the relationship between subglottal pressure and frequency. These appear to be related to biomechanical properties of the vocal folds and not to the acoustic impedance of the subglottal tract.

In the absence of a separating force, the aryepiglottic larynx narrows due to airflow. Such a downstream constriction is expected to modify the aerodynamic and acoustic load on the vocal folds.

The effect of downstream constrictions (straws) on an *in vitro* model of the vocal folds increased the intra-oral DC pressure and decreased the transglottal pressure, for the same water pressure; increased the glottal resistance, i.e. the ratio of transglottal pressure:flow; changed the calculated acoustic load from inertive to compliant in the measured range of f_0; changed the mechanical mode of oscillation in the vocal folds, from coinciding with a high mechanical resonance of the vocal fold replica, to a lower one.

It could be expected that some of these changes also occur *in vivo*, however, the non-rigid walls of the real vocal tract limit the extent to which a constriction at the lips can increase the inertance of the load 'seen' by the source.

7.3. Suggestions for Further Work

The low frequency resonance and anti-resonance of the vocal tract will have their largest phonetic effect on plosives. Making precise measurements during transients would require improved signal to noise ratios but would be an interesting extension of this work.

A detailed study of the subglottal resonances could be made with a tracheal puncture: a fine catheter linking a microphone and pressure transducer to the subglottal region. This is a highly invasive procedure but, if it were conducted for other reasons, there would be advantages in making a measurement with an impedance head at the lips.

A further series of experiments on excised larynges could employ ducts to simulate the vocal tract. This would require a suitable seal and would prohibit video imaging, but could allow direct study of phonation with and without a downstream acoustic load. Making the appropriate seal without restricting control of the geometry and tension would require some ingenuity and some flexible materials.

References

Alipour, F., Berry, D.A., Titze, I.R., 2000. A finite-element model of vocal-fold vibration. The Journal of the Acoustical Society of America 108, 3003–3012.

Alipour, F., Finnegan, E.M., Jaiswal, S., 2013. Phonatory Characteristics of the Excised Human Larynx in Comparison to Other Species. Journal of Voice 27, 441–447.

Alipour, F., Jaiswal, S., 2008. Phonatory characteristics of excised pig, sheep, and cow larynges. The Journal of the Acoustical Society of America 123, 4572. doi:10.1121/1.2908289

Alipour, F., Jaiswal, S., 2009. Glottal airflow resistance in excised pig, sheep, and cow larynges. Journal of voice : official journal of the Voice Foundation 23, 40–50. doi:10.1016/j.jvoice.2007.03.007

Alipour, F., Jaiswal, S., Finnegan, E., 2007. Aerodynamic and acoustic effects of false vocal folds and epiglottis in excised larynx models. The Annals of otology, rhinology, and laryngology 116, 135.

Alipour, F., Scherer, R.C., 2007. On pressure-frequency relations in the excised larynx. The Journal of the Acoustical Society of America 122, 2296–2305.

Baer, T., 1979. Reflex activation of laryngeal muscles by sudden induced subglottal pressure changes. The Journal of the Acoustical Society of America 65, 1271–1275.

Baer, T., Gore, J.C., Gracco, L.C., Nye, P.W., 1991. Analysis of vocal tract shape and dimensions using magnetic resonance imaging: Vowels. The Journal of the Acoustical Society of America 90, 799–828.

Bailly, L., Bernardoni, N.H., Müller, F., Rohlfs, A.-K., Hess, M., 2014. The ventricular-fold dynamics in human phonation. Journal of Speech, Language, and Hearing Research.

Bailly, L., Henrich, N., Pelorson, X., 2010. Vocal fold and ventricular fold vibration in period-doubling phonation: Physiological description and aerodynamic modeling. J Acoust Soc Am 127, 3212–3222. doi:10.1121/1.3365220

Bailly, L., Pelorson, X., Henrich, N., Ruty, N., 2008. Influence of a constriction in the near field of the vocal folds: Physical modeling and experimental validation. J Acoust Soc Am 124, 3296–3308. doi:10.1121/1.2977740

Barney, A., De Stefano, A., Henrich, N., 2007. The effect of glottal opening on the acoustic response of the vocal tract. Acta Acustica United with Acustica 93, 1046–1056.

Benade, A.H., 1985. Air column, reed, and player's windway interaction in musical instruments, in: Titze, I.R., Scherer, R.C. (Eds.), Vocal Fold Physiology, Biomechanics, Acoustics, and Phonatory Control. Denver Center for the Performing Arts, Denver, CO, pp. 425–452.

Berry, D.A., Herzel, H., Titze, I.R., Story, B.H., 1996. Bifurcations in excised larynx experiments. Journal of Voice 10, 129–138.

Birkholz, P., 2005. 3D-Artikulatorische Sprachsynthese (Phd Thesis).

Birkholz, P., 2012. VocalTractLab 2.0 User Manual.

Birkholz, P., Jackèl, D., Kroger, B., 2007. Simulation of losses due to turbulence in the time-varying vocal system. Audio, Speech, and Language Processing, IEEE Transactions on 15, 1218–1226.

Bogert, B., 1953. On the band width of vowel formants. The Journal of the Acoustical Society of America 25, 791.

Bourne, T., Garnier, M., 2012. Physiological and acoustic characteristics of the female music theater voice. The Journal of the Acoustical Society of America 131, 1586–1594.

Brown, N.J., Xuan, W., Salome, C.M., Berend, N., Hunter, M.L., Musk, A., James, A.L., King, G.G., 2010. Reference equations for respiratory system resistance and reactance in adults. Respiratory physiology & neurobiology 172, 162–168.

Campbell, D., Brown, J., 1963. The electrical analogue of lung. British Journal of Anaesthesia 35, 684–692.

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, 482–491. doi:10.1114/1.1561287

Chiche, Y., 2012. La reconstruction tridimensionnelle des structures laryngees a partir d'un modele anatomique (Master 1). Universite Joseph Fourier, Ingénierie de la Santé et du Médicament, CHU, Ingénierie de la Santé et du Médicament, Universite de Grenoble, Grenoble, France.
Childers, D.G., Diaz, J.A., 2000. Speech processing and synthesis toolboxes. The Journal of the Acoustical Society of America 108, 1975.
Childers, D.G., Wu, K., 1991. Gender recognition from speech. Part II: Fine analysis. Journal of the Acoustical Society of America 90, 1841–1856.
Chi, X., Sonderegger, M., 2004. Subglottal coupling and vowel space. The Journal of the Acoustical Society of America 115, 2540–2540.
Chi, X., Sonderegger, M., 2007. Subglottal coupling and its influence on vowel formants. The Journal of the Acoustical Society of America 122, 1735–1745.
Clark, J., Yallop, C., Fletcher, J., 2007. An introduction to phonetics and phonology, 3rd ed, Blackwell textbooks in linguistics. Blackwell Pub., Malden, MA ; Oxford.
Coffin, B., 1989. Historical vocal pedagogy classics. Scarecrow Press.
Coltman, J.W., 2003. Acoustical losses in wet instrument bores (L). The Journal of the Acoustical Society of America 114, 1221.
Cranen, B., Boves, L., 1987. On subglottal formant analysis. The Journal of the Acoustical Society of America 81, 734.
Cronjaeger, R., 1978. Die Entstehung des primären Stimmklangs im menschlichen Kehlkopf: ein Modell. na.
Dalmont, J.P., Nederveen, C., Joly, N., 2001. Radiation impedance of tubes with different flanges: Numerical and experimental investigations. Journal of Sound and Vibration 244, 505–534.
De Cheveigné, A., Kawahara, H., 2002. YIN, a fundamental frequency estimator for speech and music. The Journal of the Acoustical Society of America 111, 1917. doi:10.1121/1.1458024
Delebecque, L., Pelorson, X., Beautemps, D., Laval, X., 2013. Physical modeling of bilabial plosives production, in: Proceedings of Meetings on Acoustics. Acoustical Society of America, p. 035047.
Demolin, D., Metens, T., Soquet, A., 1996. Three-dimensional measurement of the vocal tract by MRI, in: Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on. IEEE, pp. 272–275.
Diba, C., King, G.G., Berend, N., Salome, C.M., 2011. Improved respiratory system conductance following bronchodilator predicts reduced exertional dyspnoea. Respiratory medicine 105, 1345–1351.
Dickens, P., Smith, J., Wolfe, J., 2007. Improved precision in measurements of acoustic impedance spectra using resonance-free calibration loads and controlled error distribution. J Acoust Soc Am 121, 1471–1481. doi:10.1121/1.2434764
Dunn, H., 1961. Methods of measuring vowel formant bandwidths. The Journal of the Acoustical Society of America 33, 1737.
Epps, J., Smith, J., Wolfe, J., 1997. A novel instrument to measure acoustic resonances of the vocal tract during phonation. Meas Sci Technol 8, 1112–1121.
Esling, J.H., Fraser, K.E., Harris, J.G., 2005. Glottal stop, glottalized resonants, and pharyngeals: A reinterpretation with evidence from a laryngoscopic study of Nuuchahnulth (Nootka). Journal of Phonetics 33, 383–410. doi:10.1016/j.wocn.2005.01.003
Fant, G., 1960. Acoustic theory of speech production with calculations based on X-ray studies of Russian articulations. Mouton De Gruyter.
Fant, G., 1961. The acoustics of speech, in: Proceedings of the 3rd International Congress on Acoustics. Presented at the 3rd International Congress on Acoustics (ICA), Stuttgart, pp. 188–201.
Fant, G., 1972. Vocal tract wall effects, losses, and resonance bandwidths. Speech Transmission Laboratory Quarterly progress and status report 2, 28–52.
Fant, G., 1985. The vocal tract in your pocket calculator. STL-QPSR, Royal Institute of Technology, Stockholm, Sweden 2–3.
Fant, G., 2001. On the speech code. Speech Transmission Laboratory Quarterly Progress and Status Report 42, 61–67.
Fant, G., Ishizaka, K., Lindqvist, J., Sundberg, J., 1972. Subglottal formants. STL-QPSR 1, 1–12.

Fant, G., Nord, L., Branderud, P., 1976. A note on the vocal tract wall impedance. STL-QPSR 4, 13–20.
Fant, G., Sonesson, B., 1964. Speech at high ambient air-pressure. Quarterly Progress Report, Speech Transmission Lab, 5IL-QPSR, Stockholm 9–21.
Ferrein, A., 1741. De la formation de la voix de l'homme. Mémoires de l'Académie Royale des Sciences, Paris 409–432.
Fitch, W.T., Giedd, J., 1999. Morphology and development of the human vocal tract: A study using magnetic resonance imaging. The Journal of the Acoustical Society of America 106, 1511–1522.
Flanagan, J.L., 1968. Source-system interaction in the vocal tract. Annals of the New York Academy of Sciences 155, 9–17.
Fletcher, N.H., 1993. Autonomous Vibration of Simple Pressure-Controlled Valves in Gas-Flows. J Acoust Soc Am 93, 2172–2180.
Fletcher, N.H., Rossing, T.D., 1998. The physics of musical instruments, 2nd ed. Springer, New York.
Fredberg, J.J., Hoenig, A., 1978. Mechanical response of the lungs at high frequencies. Journal of Biomechanical Engineering 100, 57–66.
Fujimura, O., Lindqvist, J., 1971. Sweep-Tone Measurements of Vocal-Tract Characteristics. The Journal of the Acoustical Society of America 49, 541.
Garnier, M., Henrich, N., Smith, J., Wolfe, J., 2010. Vocal tract adjustments in the high soprano range. Journal of the Acoustical Society of America 127, 3771–3780. doi:10.1121/1.3419907
Ghonim, A., Lim, J., Smith, J., Wolfe, J., 2013. The division of the perceptual vowel plane for different accents of English and the characteristic separation required to distinguish vowels. Acoustics Australia 41, 160–164.
Gramming, P., Sundberg, J., Ternström, S., Leanderson, R., Perkins, W.H., 1988. Relationship between changes in voice pitch and loudness. Journal of Voice 2, 118–126.
Gussenhoven, C., 2004. The phonology of tone and intonation. Cambridge University Press.
Habib, R.H., Chalker, R.B., Suki, B., Jackson, A.C., 1994. Airway geometry and wall mechanical properties estimated from subglottal input impedance in humans. Journal of Applied Physiology 77, 441–451.
Hanna, N.N., Henrich Bernadoni, N., Bennani, F., Morand, L., Pelloux, M., Mancini, A., Laval, X., Legou, T., Chaffanjon, P., 2014a. Laryngeal control of fundamental frequency in excised human larynges. Submitted.
Hanna, N.N., Henrich, N., Bennani, F., Morand, L., Pelloux, M., Legou, T., Chaffanjon, P., 2014b. Singing excised human larynges: Investigating aerodynamical and biomechanical control of phonation, in: International Conference on Voice Physiology and Biomechanics (ICVPB 2014). Presented at the International Conference on Voice Physiology and Biomechanics (ICVPB 2014), Salt Lake City, USA.
Hanna, N.N., Henrich, N., Mancini, A., Legou, T., Laval, X., Chaffanjon, P., 2013a. Singing excised human larynges: relationship between subglottal pressure and fundamental frequency. Models and analysis of vocal emissions for biomedical applications (MAVEBA) 133.
Hanna, N.N., Laval, X., Henrich, N., 2013b. Phonation into straws: impact of an aeroacoustical load on in-vitro vocal fold vibration, in: Pan European Voice Conference (PEVOC). Presented at the Pan European Voice Conference (PEVOC), Prague, Czech Republic.
Hanna, N.N., Smith, J., Wolfe, J., 2012a. Low frequency response of the vocal tract: acoustic and mechanical resonances and their losses. Presented at the Proceedings of the Australian Acoustical Society Conference.
Hanna, N.N., Smith, J., Wolfe, J., 2012b. Acoustic measurements of vocal tract resonances, in: International Conference on Voice Physiology and Biomechanics (ICVPB 2012). Presented at the International Conference on Voice Physiology and Biomechanics, (ICVPB 2012), Erlangen, Germany.
Hanna, N.N., Smith, J., Wolfe, J., 2012c. Resonances and bandwidths in the vocal tract and why they are important for speech comprehension, in: The 20th Australian Institute of Physics Congress (AIP 2012). Presented at the The 20th Australian Institute of Physics Congress (AIP 2012), Sydney, Australia.

Hanna, N.N., Smith, J., Wolfe, J., 2013c. Measurements of the aero-acoustic properties of the vocal folds and vocal tract by broad and narrow band probes during phonation into controlled acoustic loads. The Journal of the Acoustical Society of America 133, 3523–3523.
Hanson, H.M., Chuang, E.S., 1999. Glottal characteristics of male speakers: acoustic correlates and comparison with female data. The Journal of the Acoustical Society of America 106, 1064.
Harper, P., Kraman, S.S., Pasterkamp, H., Wodicka, G.R., 2001. An acoustic model of the respiratory tract. Biomedical Engineering, IEEE Transactions on 48, 543–550.
Hatzikirou, H., Fitch, W., Herzel, H., 2006. Voice instabilities due to source-tract interactions. Acta acustica united with acustica 92, 468–475.
Hawks, J.W., Miller, J.D., 1995. A formant bandwidth estimation procedure for vowel synthesis. The Journal of the Acoustical Society of America 97, 1343.
Helmholtz, H. von, Ellis, A.J., 1875. On the sensations of tone as a physiological basis for the theory of music. Longmans, Green, and Co., London,.
Henrich, N., Kiek, M., Smith, J., Wolfe, J., 2007. Resonance strategies used in Bulgarian women's singing style: A pilot study. Logopedics Phonatrics Vocology 32, 171–177.
Henrich, N., Ruty, N., 2012. What happens when a self-oscillating vocal-folds replica phonate into a straw?, in: 8th International Conference on Voice Physiology and Biomechanics (ICVPB 2012).
Henrich, N., Smith, J., Wolfe, J., 2011. Vocal tract resonances in singing: Strategies used by sopranos, altos, tenors, and baritones. The Journal of the Acoustical Society of America 129, 1024.
Hermant, N., Pelorson, X., Chouly, F., Richard, F., 2012. Finite element model of a vocal fold replica, in: 8th International Conference on Voice Physiology and Biomechanics (ICVPB 2012).
Ho, J.C., Zañartu, M., Wodicka, G.R., 2011. An anatomically based, time-domain acoustic model of the subglottal system for speech production. The Journal of the Acoustical Society of America 129, 1531–1547.
Hoppe, U., Rosanowski, F., Döllinger, M., Lohscheller, J., Schuster, M., Eysholdt, U., 2003. Glissando: laryngeal motorics and acoustics. Journal of Voice 17, 370–376.
Horáček, J., Švec, J.G., Veselý, J., Vilkman, E., 2004. Bifurcations in excised larynges caused by vocal fold elongation. Presented at the Proceedings of the International Conference on Voice Physiology and Biomechanics, pp. 87–89.
Horsfield, K., Dart, G., Olson, D.E., Filley, G.F., Cumming, G., 1971. Models of the human bronchial tree. Journal of Applied Physiology 31, 207–217.
House, A.S., Stevens, K.N., 1958. Estimation of formant band widths from measurements of transient response of the vocal tract. Journal of Speech and Hearing Research 1, 309.
Hsiao, T.-Y., Liu, C.-M., Luschei, E.S., Titze, I.R., 2001. The Effect of Cricothyroid Muscle Action on the Relation Between Subglottal Pressure and Fundamental Frequency in an< i> In Vivo Canine Model. Journal of Voice 15, 187–193.
Hudde, H., Slatky, H., 1989. The acoustical input impedance of excised human lungs—measurements and model matching. The Journal of the Acoustical Society of America 86, 475.
Ishizaka, K., Flanagan, J.L., 1972. Synthesis of Voiced Sounds From a Two-Mass Model of the Vocal Cords. Bell system technical journal 51, 1233–1268.
Ishizaka, K., French, J., Flanagan, J., 1975. Direct determination of vocal tract wall impedance. Acoustics, Speech and Signal Processing, IEEE Transactions on 23, 370–373.
Ishizaka, K., Matsudaira, M., Kaneko, T., 1976. Input acoustic-impedance measurement of the subglottal system. The Journal of the Acoustical Society of America 60, 190–197.
Jiang, J.J., Titze, I.R., 1993. A methodological study of hemilaryngeal phonation. The Laryngoscope 103, 872–882.
Johansson, C., Sundberg, J., Wilbrand, H., 1985. X-ray study of articulation and formant frequencies in two female singers, in: SMAC 83. Proceedings of the Stockholm Internat Music Acoustics Conf. pp. 203–218.
Joliveau, E., Smith, J., Wolfe, J., 2004a. Acoustics - Tuning of vocal tract resonance by sopranos. Nature 427, 116–116. doi:10.1038/427116a

Joliveau, E., Smith, J., Wolfe, J., 2004b. Vocal tract resonances in singing: The soprano voice. J Acoust Soc Am 116, 2434–2439. doi:10.1121/1.1791717
Jovicic, S.T., 1998. Formant feature differences between whispered and voiced sustained vowels. Acta Acustica United with Acustica 84, 739–743.
Kallail, K.J., Emanuel, F.W., 1984. Formant-frequency differences between isolated whispered and phonated vowel samples produced by adult female subjects. Journal of Speech, Language, and Hearing Research 27, 245–251.
Kewley-Port, D., 1983. Time-varying features as correlates of place of articulation in stop consonants. The Journal of the Acoustical Society of America 73, 322–335.
Kob, M., 2002. Physical modeling of the singing voice. Universitätsbibliothek.
Lamarche, A., Ternström, S., 2008. An exploration of skin acceleration level as a measure of phonatory function in singing. Journal of Voice 22, 10–22.
Langford-Smith, F., 1963. Radiotron Designer's Handbook, 6th ed. Amalgamated Wireless Valve Co. Pty. Ltd., Sydney, Australia.
Laukkanen, A.-M., Horáček, J., Krupa, P., Švec, J.G., 2012. The effect of phonation into a straw on the vocal tract adjustments and formant frequencies. A preliminary MRI study on a single subject completed with acoustic results. Biomedical Signal Processing and Control 7, 50–57.
Laukkanen, A.-M., Titze, I.R., Hoffman, H., Finnegan, E., 2008. Effects of a semioccluded vocal tract on laryngeal muscle activity and glottal adduction in a single female subject. Folia Phoniatrica et Logopaedica 60, 298–311.
Lieberman, P., Knudson, R., Mead, J., 1969. Determination of the rate of change of fundamental frequency with respect to subglottal air pressure during sustained phonation. The Journal of the Acoustical Society of America 45, 1537.
Lucero, J.C., Lourenço, K.G., Hermant, N., Van Hirtum, A., Pelorson, X., 2012. Effect of source–tract acoustical coupling on the oscillation onset of the vocal folds. The Journal of the Acoustical Society of America 132, 403.
Lucero, J.C., Van Hirtum, A., Ruty, N., Cisonni, J., Pelorson, X., 2009. Validation of theoretical models of phonation threshold pressure with data from a vocal fold mechanical replica (L). J Acoust Soc Am 125, 632–635. doi:10.1121/1.3056468
Lulich, S.M., 2006. The role of lower airway resonances in defining vowel feature contrasts. Massachusetts Institute of Technology.
Lulich, S.M., 2010. Subglottal resonances and distinctive features. Journal of Phonetics 38, 20–32.
Lulich, S.M., Alwan, A., Arsikere, H., Morton, J.R., Sommers, M.S., 2011. Resonances and wave propagation velocity in the subglottal airways. The Journal of the Acoustical Society of America 130, 2108. doi:10.1121/1.3632091
Lulich, S.M., Zanartu, M., Mehta, D.D., Hillman, R.E., 2009. Source-filter interaction in the opposite direction: subglottal coupling and the influence of vocal fold mechanics on vowel spectra during the closed phase. Proceedings of Meetings on Acoustics 6, 060007.
Lv, G., Zhao, H., 2009. A Modified Adaptive Algorithm for Formant Bandwidth in Whisper Conversion. IEEE, pp. 368–371.
Makarov, I., Sorokin, V., 2004. Resonances of a branched vocal tract with compliant walls. Acoustical Physics 50, 323–330.
Mau, T., Muhlestein, J., Callahan, S., Weinheimer, K.T., Chan, R.W., 2011. Phonation threshold pressure and flow in excised human larynges. The Laryngoscope 121, 1743–51. doi:10.1002/lary.21880
McCulloch, T.M., Hoffman, M.R., McAvoy, K.E., Jiang, J.J., 2013. Initial investigation of anterior approach to arytenoid adduction in excised larynges. The Laryngoscope 123, 942–7. doi:10.1002/lary.23650
Miller, D.G., Švec, J.G., Schutte, H.K., 2002. Measurement of characteristic leap interval between chest and falsetto registers. Journal of Voice 16, 8–19.
Moisik, S.R., Esling, J.H., Crevier-Buchman, L., 2010. A high-speed laryngoscopic investigation of aryepiglottic trilling. The Journal of the Acoustical Society of America 127, 1548. doi:10.1121/1.3299203
Monsen, R.B., Engebretson, A.M., 1983. The Accuracy of Formant Frequency MeasurementsA Comparison of Spectrographic Analysis and Linear Prediction. Journal of Speech, Language, and Hearing Research 26, 89–97.

Pelorson, X., Hirschberg, A., Van Hassel, R.R., Wijnands, A.P.J., Auregan, Y., 1994. Theoretical and experimental study of quasisteady-flow separation within the glottis during phonation. Application to a modified two-mass model. The Journal of the Acoustical Society of America 96, 3416–3431.

Peterson, G.E., Barney, H.L., 1952. Control methods used in a study of the vowels. Journal of the Acoustical Society of America 24, 175–184.

Pham Thi Ngoc, Y., Badin, P., 1994. Vocal tract acoustic transfer function measurements : further developments and applications. Le Journal de Physique IV 04, C5–549–C5–552. doi:10.1051/jp4:19945118

Pillot-Loiseau, C., Quattrocchi, S., Amy De La Bretèque, B., 2009. Travail de la voix sur le souffle: rééducation à la paille, aspects scientifiques et rééducatifs méthode du Dr Benoît AMY de la BRETEQUE.

Plant, R.L., Younger, R.M., 2000. The interrelationship of subglottic air pressure, fundamental frequency, and vocal intensity during speech. Journal of Voice 14, 170–177.

Rabiner, L.R., Schafer, R.W., 1978. Digital processing of speech signals. Prentice-hall Englewood Cliffs, NJ.

Robert, Y., Bennani, F., Hanna, N.N., Henrich, N., Chaffanjon, P., 2014. Construction d'un modèle laryngé hybride. Mise au point d'un banc d'essai. Morphologie 98, 136–137.

Robinson, P.D., Turner, M., Brown, N.J., Salome, C., Berend, N., Marks, G.B., King, G.G., 2011. Procedures to improve the repeatability of forced oscillation measurements in school-aged children. Respiratory physiology & neurobiology 177, 199–206.

Rothenberg, M., 1981. An interactive model for the voice source. Quarterly Prog. Status Report 1–17.

Roubeau, B., Henrich, N., Castellengo, M., 2009. Laryngeal Vibratory Mechanisms: The Notion of Vocal Register Revisited. Journal of Voice 23, 425–438. doi:10.1016/j.jvoice.2007.10.014

Russell, G.O., Cotton, J.C., 1959. Causes of good and bad voices. National Research Foundation, under a subsidy administered by the Carnegie Institute of Washington.

Ruty, N., 2007. PhD Thesis: Modeles d'interactions fluide parois dans le conduit vocal: Applications aux voix et aux pathologies.

Ruty, N., Pelorson, X., Van Hirtum, A., Lopez-Arteaga, I., Hirschberg, A., 2007. An in vitro setup to test the relevance and the accuracy of low-order vocal folds models. J Acoust Soc Am 121, 479–490. doi:10.1121/1.2384846

Savitzky, A., Golay, M.J.E., 1964. Smoothing and differentiation of data by simplified least squares procedures. Analytical chemistry 36, 1627–1639.

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, 985–995.

Silva, F., Hermant, N., Laval, X., Pelorson, X., 2014. Techniques expérimentales pour la caractérisation mécanique de maquettes in vitro de cordes vocales. Proceedings of the CFA2014.

Smith, J.R., 1995. Phasing of Harmonic Components to Optimize Measured Signal-to-Noise Ratios of Transfer-Functions. Meas Sci Technol 6, 1343–1348.

Sondhi, M.M., 1974. Model for wave propagation in a lossy vocal tract. The Journal of the Acoustical Society of America 55, 1070.

Sondhi, M.M., 1986. Resonances of a bent vocal tract. The Journal of the Acoustical Society of America 79, 1113–1116.

Stevens, K.N., 2000. Acoustic phonetics. MIT press.

Story, B.H., 1995. Physiologically-Based Speech Simulation Using AN Enhanced Wave-Reflection Model of the Vocal Tract.

Story, B.H., 2011. TubeTalker: An airway modulation model of human sound production, in: Proceedings of First International Workshop on Performative Speech and Singing Synthesis.

Story, B.H., Titze, I.R., Hoffman, E.A., 1996. Vocal tract area functions from magnetic resonance imaging. The Journal of the Acoustical Society of America 100, 537–554.

Suki, B., Habib, R.H., Jackson, A.C., 1993. Wave propagation, input impedance, and wall mechanics of the calf trachea from 16 to 1,600 Hz. Journal of Applied Physiology 75, 2755.

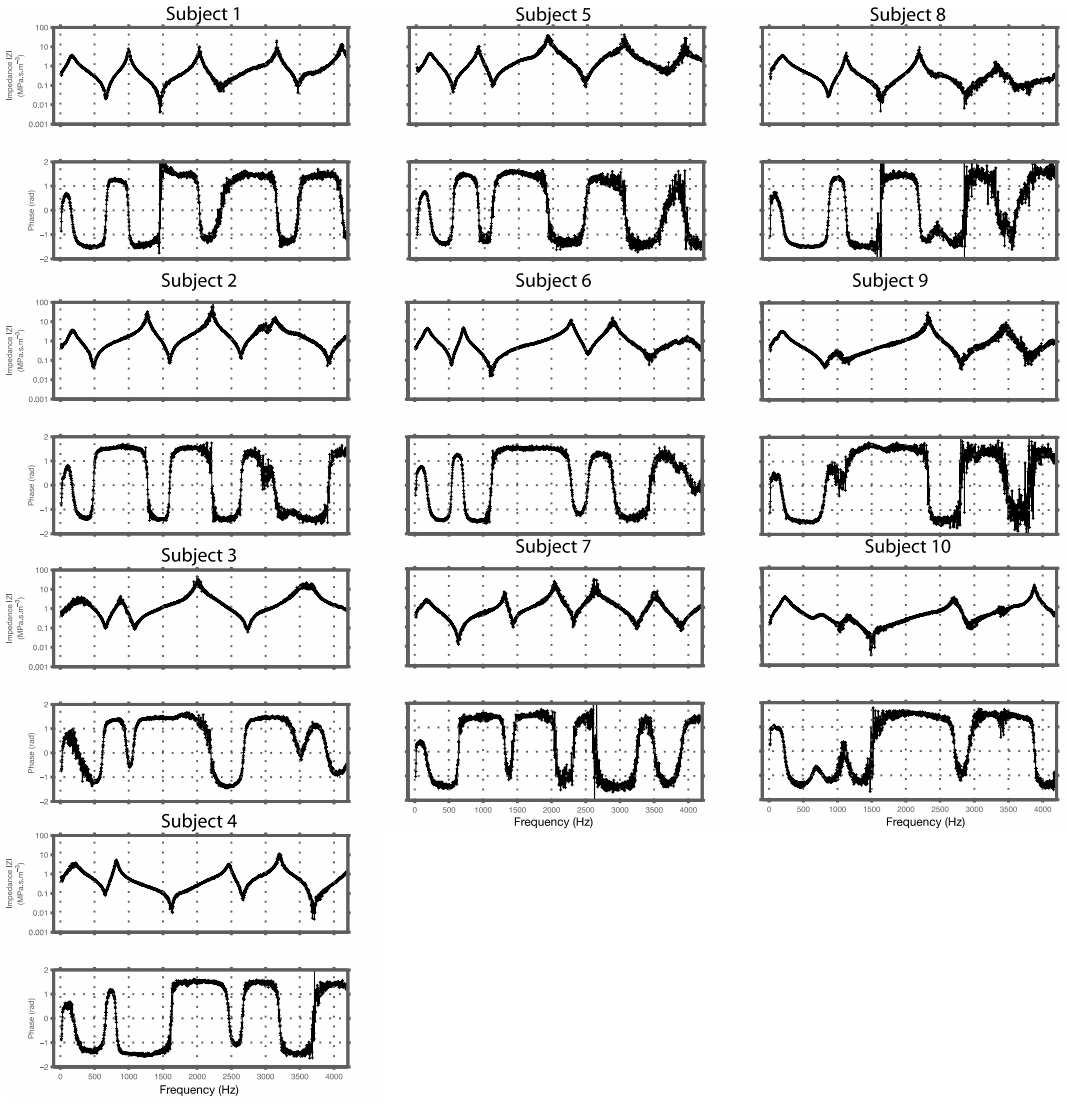
Summerfield, Q., Foster, J., Tyler, R., Bailey, P.J., 1985. Influences of formant bandwidth and auditory frequency selectivity on identification of place of articulation in stop consonants. Speech Communication 4, 213–229.
Sundberg, J., 1973. Observations on a professional soprano singer. STL-QPSR 14, 14Á.
Sundberg, J., 1973. Data on maximum speed of pitch changes. Speech transmission laboratory quarterly progress and status report 4, 39–47.
Sundberg, J., 1987. The Science of the Singing Voice (Northern Illinois UP, DeKalb, IL).
Sundberg, J., Scherer, R., Hess, M., Müller, F., Granqvist, S., 2013. Subglottal Pressure Oscillations Accompanying Phonation. Journal of Voice 27, 411–421. doi:10.1016/j.jvoice.2013.03.006
Sundberg, J., Titze, I., Scherer, R., 1993. Phonatory control in male singing: A study of the effects of subglottal pressure, fundamental frequency, and mode of phonation on the voice source. Journal of Voice 7, 15–29.
Švec, J.G., Horáček, J., Šram, F., Veselỳ, J., 2000. Resonance properties of the vocal folds: in vivo laryngoscopic investigation of the externally excited laryngeal vibrations. The Journal of the Acoustical Society of America 108, 1397–1407.
Švec, J.G., Schutte, H.K., Miller, D.G., 1996. A subharmonic vibratory pattern in normal vocal folds. Journal of speech and hearing research 39, 135–143.
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. The Journal of the Acoustical Society of America 106, 1523–1531.
Swerdlin, Y., Smith, J., Wolfe, J., 2010. The effect of whisper and creak vocal mechanisms on vocal tract resonances. J Acoust Soc Am 127, 2590–2598. doi:10.1121/1.3316288
Tarnóczy, T.H., 1962. Vowel formant bandwidths and synthetic vowels. The Journal of the Acoustical Society of America 34, 859.
Tarnopolsky, A.Z., Fletcher, N.H., Lai, J.C.S., 2000. Oscillating reed valves - An experimental study. J Acoust Soc Am 108, 400–406.
Titze, I., 2002. How to use the flow-resistant straws. J Singing 58, 429–30.
Titze, I.R., 1974. The Human Vocal Cords: A Mathematical Model. Phonetica 29, 1–21.
Titze, I.R., 1989. On the relation between subglottal pressure and fundamental frequency in phonation. The Journal of the Acoustical Society of America 85, 901.
Titze, I.R., 1994. Principles of voice production. Prentice Hall, Englewood Cliffs, N.J.
Titze, I.R., 2004. A theoretical study of f0-f1 interaction with application to resonant speaking and singing voice. Journal of Voice 18, 292–298. doi:10.1016/j.jvoice.2003.12.010
Titze, I.R., 2004. Theory of glottal airflow and source-filter interaction in speaking and singing. Acta Acust United Ac 90, 641–648.
Titze, I.R., 2006. Voice training and therapy with a semi-occluded vocal tract: rationale and scientific underpinnings. Journal of Speech, Language, and Hearing Research 49, 448.
Titze, I.R., 2008. Nonlinear source-filter coupling in phonation: Theory. J Acoust Soc Am 123, 2733–2749. doi:10.1121/1.2832337
Titze, I.R., 2009. Phonation threshold pressure measurement with a semi-occluded vocal tract. Journal of Speech, Language, and Hearing Research 52, 1062–1072.
Titze, I.R., 2011. Vocal fold mass is not a useful quantity for describing F0 in vocalization. Journal of Speech, Language, and Hearing Research 54, 520–522.
Titze, I.R., Hunter, E.J., 2011. Feasibility of measurement of a voice range profile with a semi-occluded vocal tract. Logopedics Phoniatrics Vocology 36, 32–39.
Titze, I., Riede, T., Popolo, P., 2008. Nonlinear source–filter coupling in phonation: Vocal exercises. The Journal of the Acoustical Society of America 123, 1902.
Titze, I.R., Story, B.H., 1997. Acoustic interactions of the voice source with the lower vocal tract. J Acoust Soc Am 101, 2234–2243.
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. The Journal of the Acoustical Society of America 122, 519. doi:10.1121/1.2741210
Traser, L., Burdumy, M., Richter, B., Vicari, M., Echternach, M., 2013. The Effect of Supine and Upright Position on Vocal Tract Configurations During Singing—A Comparative Study in Professional Tenors. Journal of Voice 27, 141–148. doi:10.1016/j.jvoice.2012.11.002

Valdes, J., 2013. Adaptation de clones orofaciaux à la morphologie et aux stratégies de contrôle de locuteurs cibles pour l'articulation de la parole. Institut National Polytechnique de Grenoble-INPG.
Vallabha, G.K., Tuller, B., 2002. Systematic errors in the formant analysis of steady-state vowels. Speech Communication 38, 141–160.
Vampola, T., Laukkanen, A.M., Horacek, J., Svec, J.G., 2011. Vocal tract changes caused by phonation into a tube: a case study using computer tomography and finite-element modeling. The Journal of the Acoustical Society of America 129, 310–5. doi:10.1121/1.3506347
Van den Berg, J., 1955. Transmission of the vocal cavities. The Journal of the Acoustical Society of America 27, 161.
Van den Berg, J., 1958. Myoelastic-aerodynamic theory of voice production. Journal of Speech, Language, and Hearing Research 1, 227–244.
Van den Berg, J., 1960. An electrical analogue of the trachea, lungs and tissues. Acta Physiol. Pharmacol. Neerlandica 9, 361–385.
Van den Berg, J., Tan, T., 1959. Results of experiments with human larynxes. ORL 21, 425–450.
Van den Berg, J.W., Zantema, J.T., Doornenbal Jr, P., 1957. On the air resistance and the Bernoulli effect of the human larynx. The Journal of the Acoustical Society of America 29, 626–631.
Van Hirtum, A., Cisonni, J., Ruty, N., Pelorson, X., Lopez, I., van Uittert, F., 2007. Experimental validation of some issues in lip and vocal fold physical models. Acta Acust United Ac 93, 314–323.
Vilain, C.E., Pelorson, X., Hirschberg, A., Le Marrec, L., Op't Root, W., Willems, J., 2003. Contribution to the physical modeling of the lips. Influence of the mechanical boundary conditions. Acta Acustica united with Acustica 89, 882–887.
Wakita, H., Fant, G., 1978. Toward a better vocal tract model. STL-QPSR, Royal Institute of Technology, Stockholm, Sweden 1, 9–29.
Wang, S., Alwan, A., Lulich, S.M., 2008. Speaker normalization based on subglottal resonances. Presented at the Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on, IEEE, pp. 4277–4280.
Weibel, E.R., 1963. Geometry and dimensions of airways of conductive and transitory zones. Springer.
Wolfe, J., n.d. Formant: what is a formant? http://www.phys.unsw. edu.au/jw/formant.html.
Wolfe, J., Almeida, A., Chen, J.M., George, D., Hanna, N.N., Smith, J., 2013. The player-wind instrument interaction.
Wolfe, J., Garnier, M., Smith, J., 2009a. Voice Acoustics: an introduction.
Wolfe, J., Garnier, M., Smith, J., 2009b. Vocal tract resonances in speech, singing, and playing musical instruments. Hfsp J 3, 6–23. doi:10.2976/1.2998482
Wollock, J., 1997. The noblest animate motion: Speech, physiology and medicine in pre-Cartesian linguistic thought. John Benjamins Publishing.
Xu, Y., Sun, X., 2002. Maximum speed of pitch change and how it may relate to speech. The Journal of the Acoustical Society of America 111, 1399. doi:10.1121/1.1445789
Zañartu, M., Ho, J.C., Mehta, D.D., Hillman, R.E., Wodicka, G.R., 2013. Subglottal impedance-based inverse filtering of voiced sounds using neck surface acceleration. Audio, Speech, and Language Processing, IEEE Transactions on 21, 1929–1939.
Zhang, Z., Neubauer, J., Berry, D.A., 2006. The influence of subglottal acoustics on laboratory models of phonation. The Journal of the Acoustical Society of America 120, 1558. doi:10.1121/1.2225682

Appendix A: Supporting Information

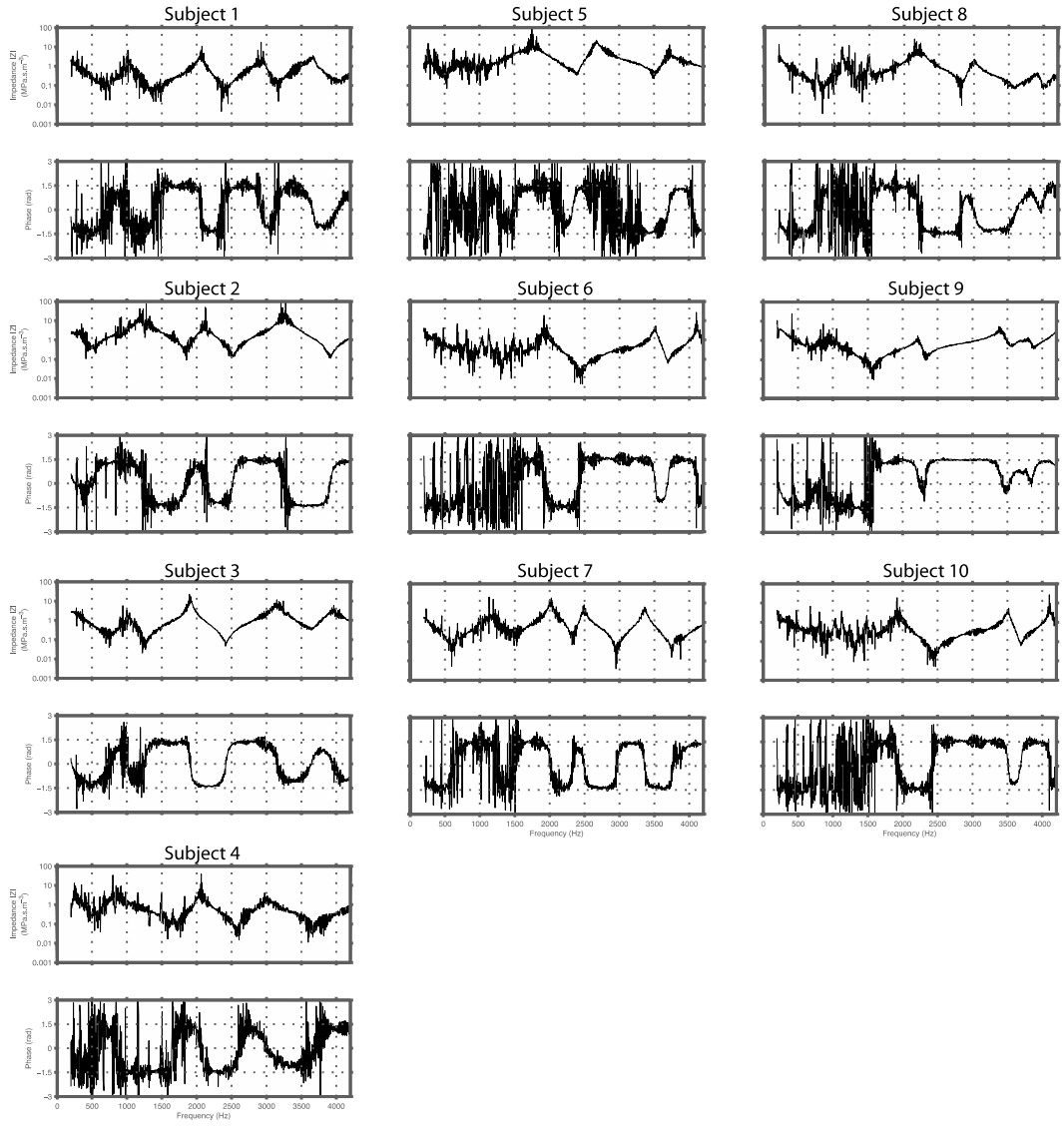
Closed glottis: impedance measurements

The graphs show example impedance magnitude and phase plots for male subjects 1-7 and female subjects 8-10. Each plot shows one 370 ms cycle measurement over the range 14-4200 Hz.



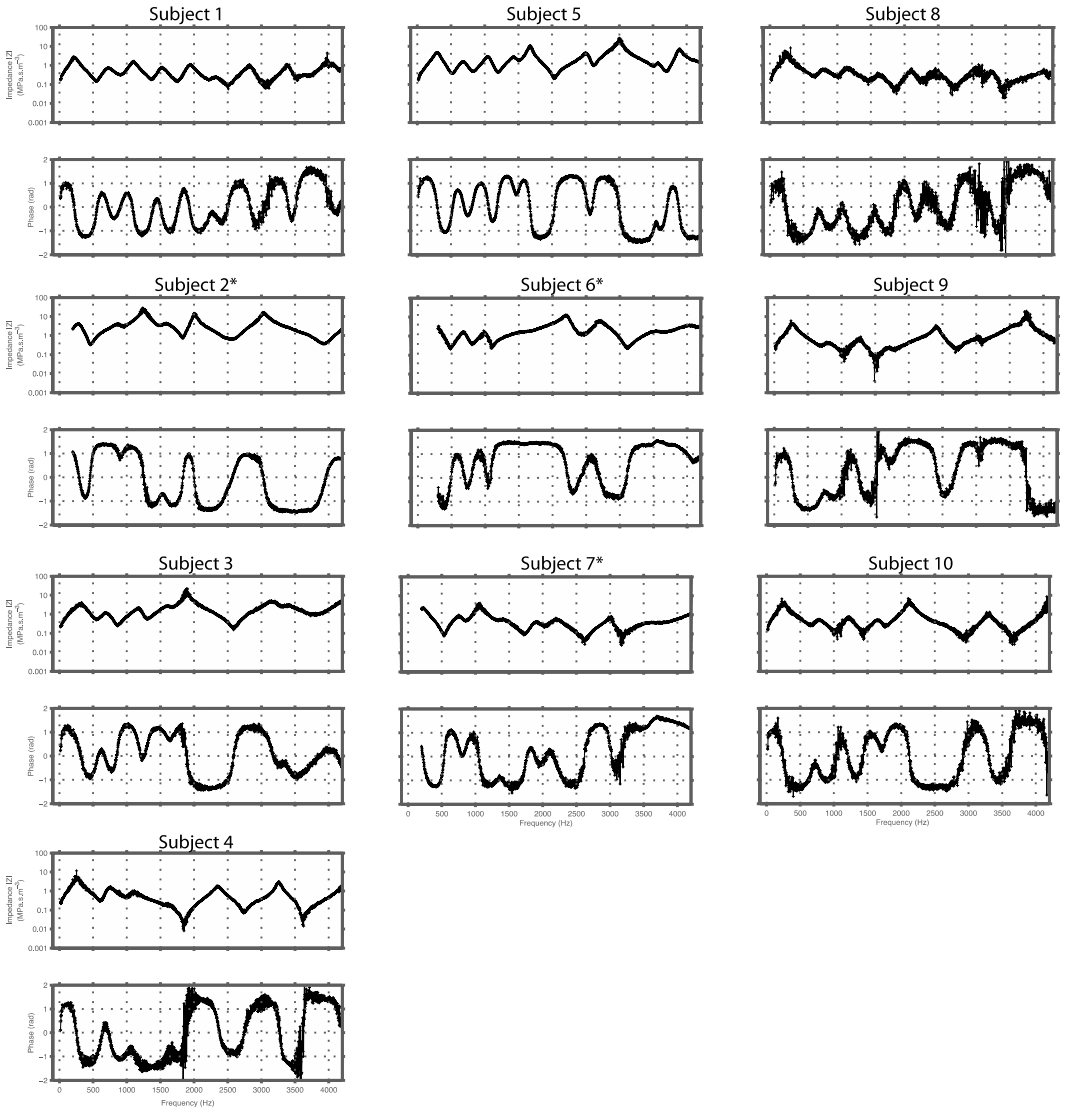
Phonation: impedance measurements

The graphs below show example impedance magnitude and phase plots for male subjects 1-7 and female subjects 8-10, phonating in low pitched, modal voice. Each plot shows one 370 ms cycle measurement over the range 14-4200 Hz.



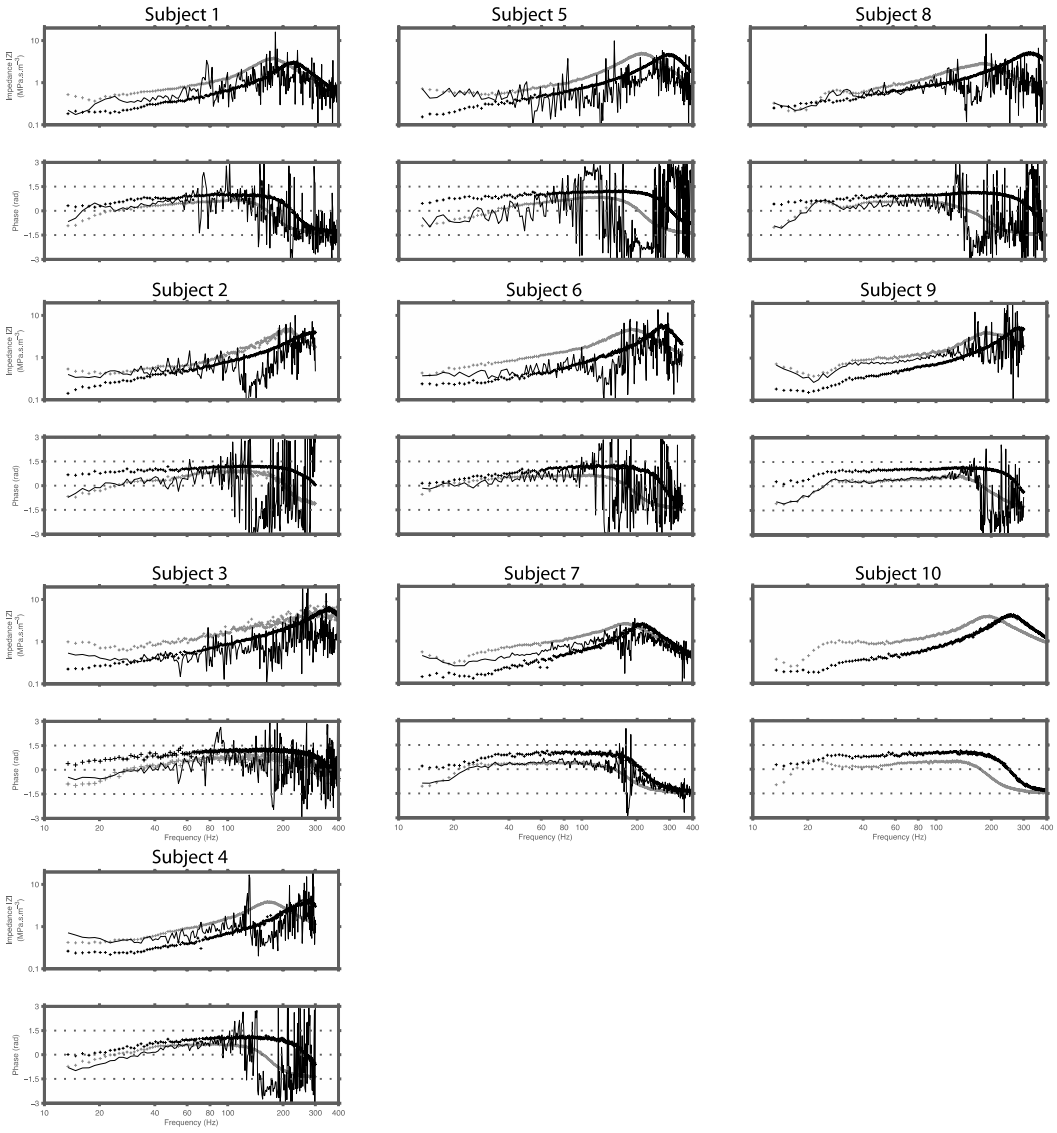
Inhalation: impedance measurements

The graphs below show example impedance magnitude and phase plots for male subjects 1-7 and female subjects 8-10. Each plot shows one 370 ms cycle measurement over the range 14-4200 Hz, except for the subjects marked with an asterisk, for whom the low frequency limit was 200 or 300 Hz to improve signal:noise ratio.



Low frequency: impedance measurements

The graphs below show example impedance magnitude and phase plots for male subjects 1-7 and female subjects 8-10. Each plot shows one measurement over the range 14-300 or 14-400 Hz. The pale markers show measurements with glottis closed, the solid line shows phonation, and the dark markers show inhalation.



Appendix B: Reprints of Articles

Low frequency response of the vocal tract: acoustic and mechanical resonances and their losses

Noel Hanna (1,2), John Smith (1) and Joe Wolfe (1)

(1) School of Physics, The University of New South Wales, Sydney, 2052, Australia

(2) Department of Speech and Cognition, GIPSA-lab, Université de Grenoble, France

ABSTRACT

The impedance spectrum of the vocal tract was measured at the lips from 10 Hz to 4.2 kHz using the three-microphone, three-calibration technique. A broadband signal synthesised from sine waves allows high precision measurements irrespective of the fundamental frequency of phonation. From these measurements, the frequencies, magnitudes and bandwidths of the resonances and antiresonances are determined directly. Resonances have impedance magnitudes of 20-100 kPa.s.m^{-3} and antiresonances 2-10 MPa.s.m^{-3}. The bandwidths measured with a closed glottis are typically around 50 Hz for resonances and antiresonances between 400 Hz and 2 kHz and increase slightly outside this range. This is in qualitative agreement with previous measurements estimated from phonations measured outside the mouth. The measured bandwidths are discussed in relation to viscothermal losses at the duct boundary, radiation from the mouth and mechanical losses in the surrounding tissues.

BACKGROUND

The acoustic features that distinguish speech phonemes are largely determined by resonances of the vocal tract. These resonance frequencies, *F*i, and how they vary, characterise vowels and are important to many consonants (Fant 1970). The bandwidths of these resonances also provide valuable information for listeners (Summerfield *et al*. 1985, Brown *et al*. 2010), for speaker identification (Childers and Diaz 2000), and provide information about the glottal source (Childers and Wu 1991, Flanagan 1972, Rabiner and Schafer 1978, and Rothenberg 1981). Further, these resonances in singers can be adjusted to boost the radiation of the fundamental and/or one of the harmonics of the voice (resonance tuning; see Joliveau *et al*. 2004a, 2004b, Henrich *et al*. 2011).

At a fundamental level, the magnitude of the bandwidths is vital to speech communication. A large bandwidth entails a low quality factor or *Q*, so a resonance with a very wide bandwidth would create a weak formant (peak in the spectral envelope): it would give only modest boost to frequencies lying nearby and so would render vowels and some other phonemes difficult or impossible to identify. A resonance with a high *Q* and a very narrow bandwidth would give a large boost to a voice harmonic that fell close to the resonant frequency. However, the harmonics of the voice would usually not fall within a very narrow bandwidth. For a voice with fundamental frequency 150 Hz, and therefore a harmonic spacing of 150 Hz, the chance that a harmonic would fall in a 10 Hz bandwidth is only about 1/15. We can argue in the reverse direction: that we can recognise phonemes characterised by formants means that the bandwidth must fall in the 'Goldilocks range': large enough to include at least one harmonic in most phonations, but narrow enough to give that harmonic a significant boost. Despite this importance to speech production and perception, we know of no previous direct measurements of the bandwidths of tract resonances during phonation.

Previous studies of the vocal tract resonances have used either the voice sound itself, swept sine excitation at the neck with the glottis closed (Peterson and Barney 1952, Van den Berg 1955, Dunn 1961, Fant 1972, Fujimura and Lindqvist 1971) or a broadband sound or vibration injected outside the neck (Pham Thi Ngoc and Badin 1994) or at the open mouth (Epps *et al*. 1997).

Using the voice restricts the frequency resolution to a value comparable with the fundamental frequency and makes it difficult to estimate bandwidths. Measurements made with the sound source at the neck are subject to an unknown transfer function. As for the broadband technique at the mouth, the low magnitude of the radiation impedance at the mouth makes it difficult to measure the magnitude of the impedance associated with the resonances and thus makes it difficult to calculate the bandwidths.

The technique used for the measurements in this study is impedance measurement at the mouth, with the lips sealed around the impedance head. It suffers from none of the disadvantages listed above. It does however have the disadvantage that sealing the lips around the impedance head (diameter 26 mm) effectively fixes the position of the first antiresonance. Further, it measures from the mouth end, whereas a measurement from the glottis end would be more immediately useful in many applications.

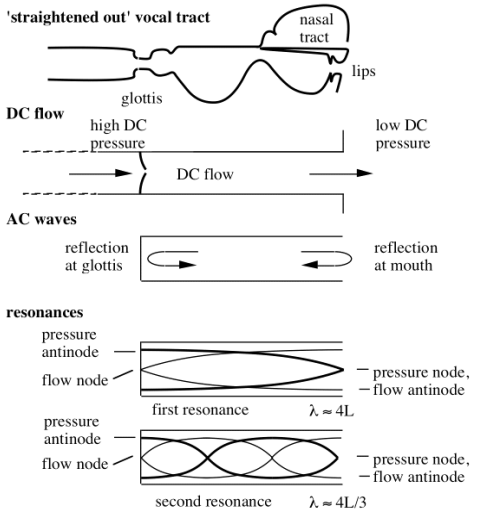


Figure 1. Comparison of a straightened vocal tract with a hard walled pipe (from Wolfe *et al*. 2009).

Figure 1 shows a considerable oversimplification of the tract geometry. For the frequencies of interest, wavelengths are much greater than the transverse dimensions, so a one-dimensional approximation captures much of the physics. In Figure 1 the vocal tract is first shown 'straightened out', which has little effect at the frequencies of interest (Sondhi 1986). Next it is imagined as a cylinder with rigid walls. For the vowel sound in the word 'heard' (denoted [əː] in the International Phonetic Alphabet), the acoustic properties over the range 300-3000 Hz are moderately well modelled by such a cylinder. At higher frequencies, the shape of the tract on the cm scale becomes more important, and at lower frequencies, the assumption of rigid wall is less justified (Sondhi 1974). We return to this comparison in the Discussion.

METHOD

Three-microphone three-calibration technique

Precise measurements of vocal tract resonances are made during normal phonation by injecting a broadband sound signal at the lips, using the three-microphone, three-calibration technique developed at UNSW (Dickens *et al*. 2007). The subjects' lips were sealed around the measurement system and the impedance spectrum was measured from 14 Hz to 4.2 kHz using a signal synthesised from sine waves (Smith 1995) spaced at 2.69 Hz (44.1 kHz/2^{14}). One complete cycle of this broadband signal will last some 370 ms. A narrow tube in parallel with the loudspeaker allows a DC airflow so measurements can be made during phonation.

The upper frequency limit of measurements is determined by the smallest spacing between two of the three microphones, which here causes a singularity at 4.3 kHz. The lower frequency limit depends on the loudspeaker, which outputs comparatively little power at low frequencies, and the microphones (B&K 4944A), which have significantly reduced sensitivity below 10 Hz.

The power of the measurement signal is distributed over hundreds of different frequencies, while the power of the voice is concentrated on a relatively small number of harmonics. This limits the possible signal:noise ratio. Thus, while it is possible to cover a very wide frequency range (e.g. 10-4.2 kHz) with a single measurement, even during phonation, using narrower frequency ranges improves the signal/noise ratio and results in better measurements from shorter time periods, particularly when loud phonation is involved.

The presence of the injected low frequency sound can excite mechanical vibration of the tissues surrounding the vocal tract at low frequencies. As we explain below, there is a mechanical resonance at ~20 Hz. Strong excitation of this resonance can perturb normal phonation, adding a strong vibrato that cannot be controlled by the subject, and that can also alter the vocal tract resonances in some cases. For this reason, we made measurements covering four frequency ranges: two to characterise the low frequency mechanical resonances (below 300 Hz), another for the higher frequency acoustic resonances (above 200 Hz) and a fourth covering the whole frequency range.

The lowest frequency feature is the first minimum in impedance, due to the mechanical resonance of the cheeks and the tissue under the jaw. This was studied using a broadband signal over the frequency range of 10-50 Hz and a frequency spacing of 0.34 Hz (44.1 kHz/2^{17}). The first maximum in impedance also involves tissue motion, discussed below. To study both the first maximum and minimum, a range of 14-300 or 400 Hz and resolution of 0.67 Hz (44.1 kHz/2^{16}) was used. Finally, for the acoustic resonances that correspond to the formants of speech ($R1$-$R5$), a frequency range of 200-4200 Hz with a resolution of 2.69 Hz (44.1 kHz/2^{14}) was used.

The subjects (seven male, three female) were asked to find a comfortable position for their mouths ensuring an air-tight seal with their lips around the impedance head. With the mouth open and the tongue low in the mouth the sound corresponded approximately to the vowel [əː]. Five of the subjects placed their teeth outside the head during all measurements, while the other five did not. Because the mouth is sealed the frequency of the first resonance is effectively fixed, so by keeping the tongue position constant (tongue low) repeatable measurements that are comparable across subjects were possible. Any effect caused by different positions of the teeth would be evident at higher frequencies that are not important for this study.

During the low frequency measurements (10-50 Hz and 14-400 Hz), four of the male subjects had a small magnet attached to their cheek and/or neck in order to measure the velocity of the tissue by recording the induced EMF in a coil of wire at a fixed distance from the magnet. The voltage-velocity relation for the magnet and coil was calibrated by oscillating the magnet with a shaker and measuring simultaneously the vibration amplitude and the induced EMF in the coil.

RESULTS AND DISCUSSION

Vocal tract model

Before presenting results, we continue with the simple model of Figure 1 to allow qualitative comparison with measurements. First, consider a cylinder with length L = 160 mm and diameter 30 mm, closed at the far (glottis) end, and with rigid walls. The calculated impedance spectrum (magnitude and phase) is shown in Figure 2 (pale dotted line). Maxima occur when $L = \lambda/2, 2\lambda/2, 3\lambda/2$ etc, *i.e*. near 1, 2, 3 kHz etc. Minima occur when $L = \lambda/4, 3\lambda/4, 5\lambda/4$ etc.

At very low frequencies, the impedance of a closed, entirely rigid cylinder is very large: with DC flow, the pressure would rise until the walls ruptured. The vocal tract is not rigid, of course. To model this in a very simple way, we consider the tract to be an acoustically compact object at low frequencies where its dimensions are $<< \lambda$, so the pressure is uniform within the tract. The walls can be simply modelled as having a mass m that is accelerated by the (approximately uniform) low frequency pressure in the tract. This mass is supported on tissues that we model as having a spring constant k and a linear loss coefficient R. If the air pressure acts on an area A of tissue, the tissue surrounding the tract has an acoustic inertance $L_t = m/A^2$ and an acoustic compliance $C_t = A^2/k$. The low frequency model of the tract now comprises these elements, plus the compliance C of the tract, treated as a compact object at these frequencies. The equivalent circuit is sketched in the inset to Figure 2.

At these low frequencies, the air in the tract is approximately a compliance in parallel with the tissue, because pressure in the tract causes displacements both by displacing the tissue and compressing the air in the tract. The calculated response of the closed cylinder with walls having parameters m, k and R as described above is also shown in Figure 2 (pale dashed line). There is a new low frequency minimum, corresponding to the series resonance of L_t and C_t, and above that, a new

low frequency maximum, corresponding to the parallel resonance of L_t and C.

Low frequency behaviour of the vocal tract

Figure 3 shows measurements of the impedance of the tract made, first with the glottis closed (*cf*. closed tube) and no phonation, then during phonation. Note the qualitative similarity to the dashed line in Figure 2.

The roughly equal spacing of the acoustic maxima and minima are explained by noting that, with the mouth aperture fixed around the measurement head, and the tongue low in the mouth, the configuration is close to that of the neutral vowel [əː], which is known to have roughly equally spaced formants and so is the vowel that best approximates a simple cylinder. Departures from equal spacing at very high frequencies are probably due to the non-uniform cross-sectional area of the vocal tract and departures from cylindrical shape on the cm scale.

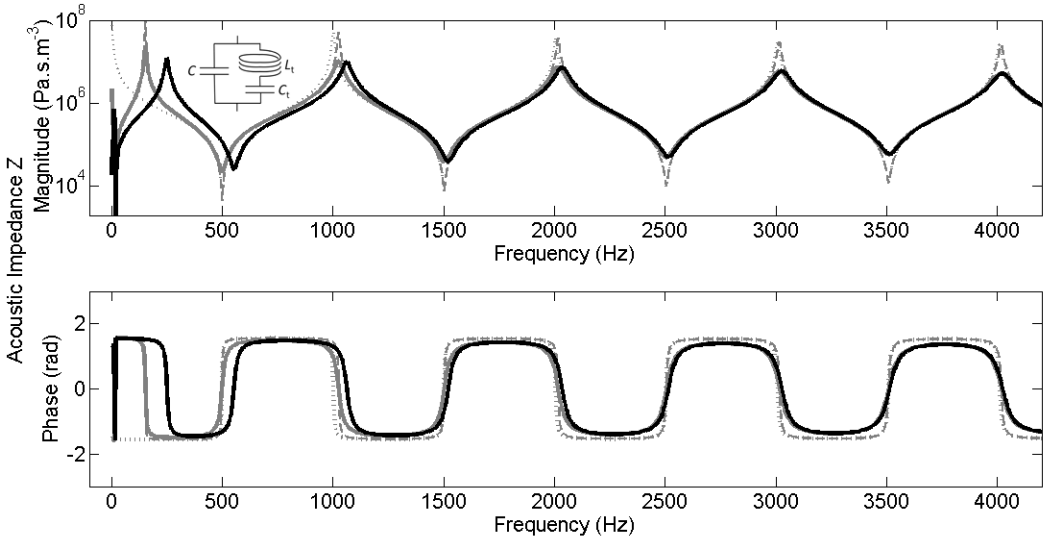


Figure 2. Theoretical impedance (magnitude and phase) of the simple models. Dotted pale line: a rigid-walled, closed cylindrical tube. Pale dashed line: cylindrical tube with finite wall mass and compliance. Pale continuous: a cylindrical tube with finite wall mass and compliance and a loss factor five times greater than the visco-thermal losses for a rigid tube. Dark continuous: as previously, but with the glottis represented by an opening area 32 mm^2, length 14 mm in accordance with the values from (Hoppe *et al*. 2003). The inset shows an electrical equivalent circuit of the low frequency model where the air in the tract is a compact compliance C and the tissue surrounding the tract has an acoustic inertance $L_t = m/A^2$ and an acoustic compliance $C_t = A^2/k$.

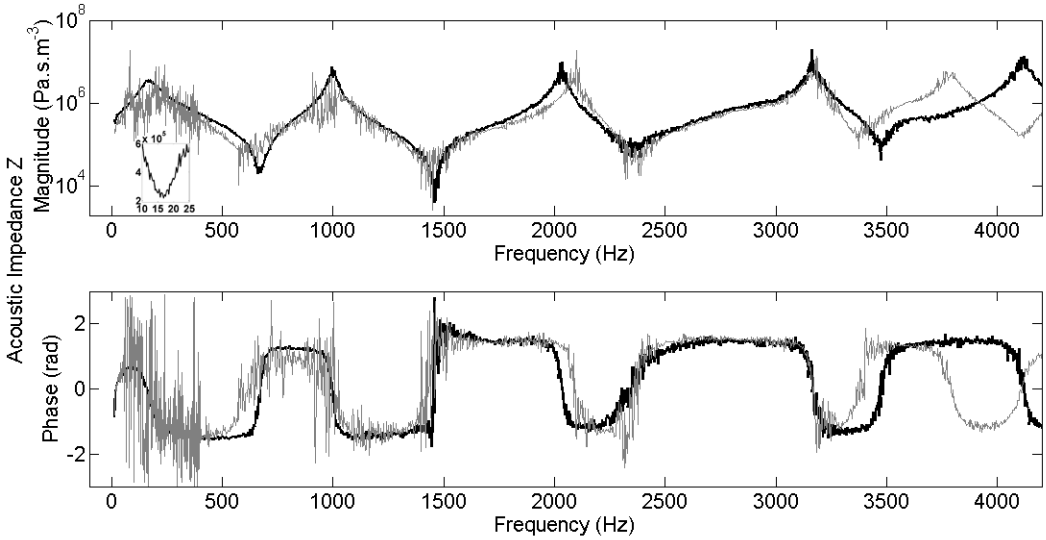


Figure 3. Measured impedance (magnitude and phase) taken from a single cycle of the broadband signal for one subject with closed glottis (dark), and phonating in mechanism 1 (pale). The data from two or more such cycles measured during the same phonation are used to produce the mean and standard deviations shown in Figure 4 and Figure 5. The inset on the upper plot shows a linear plot of the frequency range around the first minimum.

The solid lines in Figure 2 have a larger loss factor α, as described in the Discussion, to produce a curve whose extrema have magnitudes comparable with those in Figure 3.

The plots in Figure 3 are taken from measurements of one subject during only a single cycle of the broadband signal lasting 0.37 s. The pale line measurement displays noise produced by the air flow associated with phonation. Subsequent data analysis is performed on the mean of several such measurements during the same phonation gesture.

As expected, the finite mass and rigidity of the tissues surrounding the tract give rise to the extra minimum and maximum at low frequency: the first minimum is 300-500 kPa.s.m^{-3} at ~20 Hz due to the series resonance of the tissue oscillating on its own elasticity; the first maximum is 3-4 MPa.s.m^{-3} at ~200 Hz, which we attribute to the same tissue oscillating on the 'spring' of the enclosed air.

In all our measurements, the ratio of the frequencies of the first minimum and maximum was about one order of magnitude. This allows very simple modelling to estimate the tissue inertance and compliance L_t and C_t. First, we note that the compact compliance of the air in the tract, C, is determined by the volume of the vocal tract. C is approximately an open circuit at low frequency to a reasonable approximation, so the first minimum occurs at the series resonance

$$f_{0\min} \approx 20Hz \sim \frac{1}{2\pi\sqrt{L_t C_t}} \qquad (1)$$

At high frequency, C_t is approximately a short circuit, so the first maximum occurs at the parallel resonance

$$f_{0\max} \approx 200Hz = \frac{1}{2\pi\sqrt{L_t C}} \qquad (2)$$

Taking a volume of V~1x10^{-4} m^3 as an approximate value for the tract, these two equations yield L_t and C_t from any closed glottis measurement.

The low frequency compliance of the air in the vocal tract can be estimated from standard atmospheric conditions, where γ is the adiabatic constant and P_A is atmospheric pressure, so

$$C = V/\gamma P_A \sim 7\text{x}10^{-10}\ \text{m}^3\text{Pa}^{-1} \qquad (3)$$

Substituting C into Equation (2) gives a tissue inertance L_t of ~ 900 kg.m^{-4}. This can be used to give an estimate of the thickess of tissue, w, involved in the vibration. Assuming that the vocal tract is cylindrical we can take its inner surface area A as of the order 0.01 m^2. This is very roughly the area of the tissues (cheeks and under the jaw) that is observed to vibrate. The density ρ of the tissue surrounding the tract is about 10^3 kg.m^{-3}, then Equation (4) gives a plausible tissue thickness w of ~0.01 m.

$$L_t = \frac{m}{A^2} = \frac{\rho V}{A^2} = \frac{\rho w}{A} \qquad (4)$$

The compliance of the tissue can now be determined by substituting L_t into equation (1), giving C_t ~ 7x10^{-8} m^{3}Pa^{-1}. The spring constant of the tissue k can be determined by considering

$$C_t = \frac{\mathrm{d}V}{\mathrm{d}P} = \frac{A\mathrm{d}x}{\mathrm{d}F/A} = \frac{A^2}{k} \qquad (5)$$

Again assuming an area A = 0.01 m^2, the calculated spring constant is k ~ 1000 Nm^{-1}, which appears reasonable for biological tissue. A summary of these values is given in Table 1.

The amplitudes and distributions of vibrations in the cheeks and neck are not reported here but are consistent with the pattern mapped out by Fant *et al*. (1976).

Bandwidth and Q factor

In order to estimate the bandwidth and quality factor Q of the resonances, a Savitzky–Golay smoothing filter (Savitzky and Golay 1964) was used to identify the maxima and minima of the magnitude of acoustic impedance. A parabola was then fitted locally to the data near these maxima and minima. The bandwidth was determined as the frequency range below $1/\sqrt{2}$ times the maximum impedance for antiresonances or $\sqrt{2}$ times the minimum impedance for resonances.

As it is difficult for subjects to maintain exactly the same gesture over several seconds without moving, averages were taken over two or more cycles, during which the impedance magnitude curves were similar. The mean bandwidths calculated from such averages for a single subject are shown in Figure 4.

For this and other measurements with closed glottis, the bandwidths for the acoustic resonances (those above 250 Hz) are approximately 50 Hz, with a slight increase at higher frequencies. Such an increase is expected due to the increase with frequency of the visco-thermal losses at the boundary layer with the walls of a tube. A similar weak dependance on frequency can also be seen in the decreasing magnitude and increasing bandwidths of the first two theoretical curves in Figure 2.

	Tract	Tissue		
Acoustic	C = 7x10^{-10} m^{3}Pa^{-1}	L_t = 900 kg.m^{-4}	C_t = 7x10^{-8} m^{3}Pa^{-1}	(Fitted to Data)
Geometric	V = 10^{-4} m^3	A = 0.01 m^2		(Assumed)
		w = 0.01 m	k = 1000 Nm^{-1}	(Calculated)

Table 1. Low frequency model resonance parameters

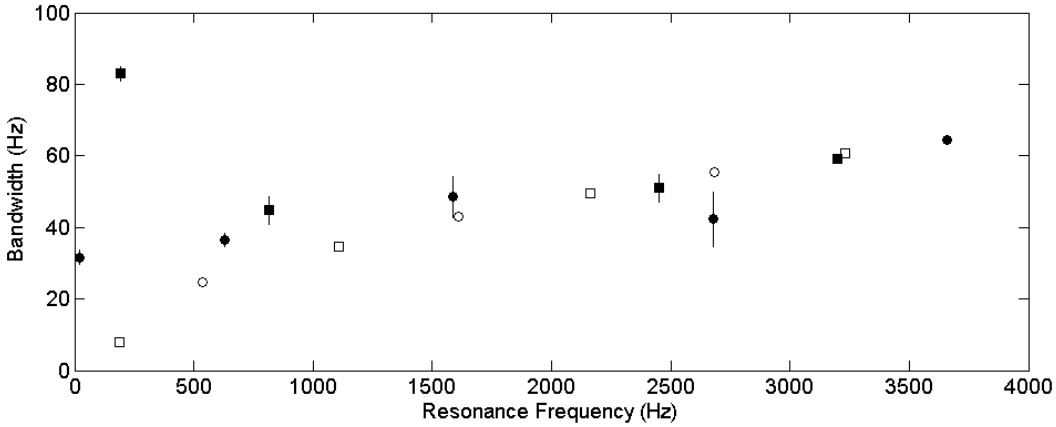


Figure 4. Mean measured bandwidth as a function of resonance frequency for one subject with glottis closed. Closed circles indicate impedance minima, closed squares indicate impedance maxima. Data are for multiple measurements and error bars show ±1 standard deviation. The open shapes show the bandwidths of the cylindrical tube with finite wall mass and compliance and a loss factor α five times greater than the visco-thermal losses for a rigid tube.

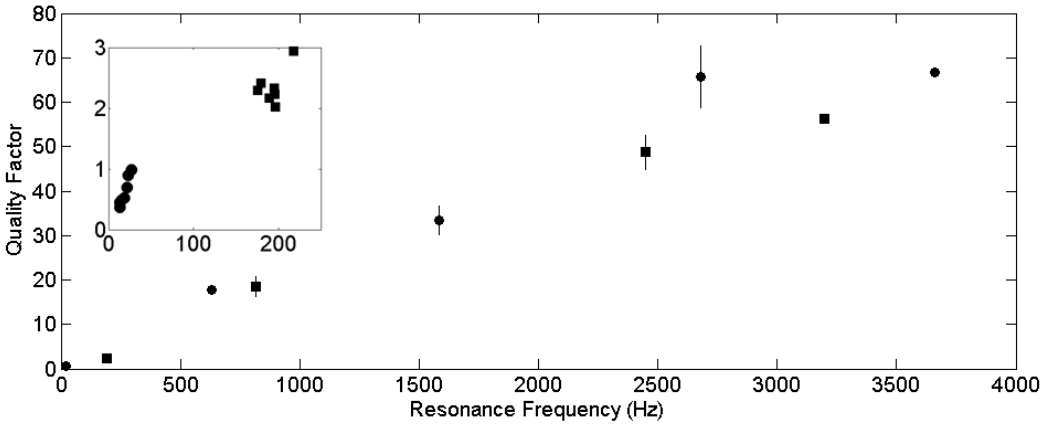


Figure 5. Mean Q factor as a function of resonance frequency measured for one subject with glottis closed. Circles indicate impedance minima, squares indicate impedance maxima, error bars show ±1 standard deviation. Inset: an expansion for the first impedance minimum and maximum.

Figure 5 shows the Q factor (ratio of frequency to bandwidth) as a function of resonance frequency for the same measurements as Figure 4. The Q factor for the acoustic bandwidths shows a dependance on the frequency consistent with losses that increase with increasing frequency.

Because these measurements were made with the impedance head sealed by the subjects' lips, the bandwidths shown do not include radiation loss from the open mouth. When the mouth is opened, the boundary condition at the mouth—the radiation impedance at the mouth—is largely inertive, but it also includes a real, resistive term. Because of transmission and hence loss of energy at the open end, the impedance spectrum of an open cylinder shows impedance extrema whose magnitudes decrease more strongly with frequency than do those of a closed cylinder. So the next step is to use the losses obtained from these experiments to estimate the bandwidths that would apply to the vocal tract with the mouth open.

In this preliminary study, the method used is simply to find the value of the (complex) visco-thermal loss factor α that, when applied to a cylindrical closed tube, gives an impedance spectrum with extrema values similar to those measured here. An increase in α by a factor of five, shown in Figure 4, was required to produce the solid line plots in Figure 2, whose extrema are similar to those of the experimental measurement in Figure 3. To model the input impedance of the tract as 'seen' from the glottis, we use this factor of α to calculate the input impedance at one end of a cylinder that is open and flanged at the far end. This gives an estimation of the impedance at the glottis, which is shown in Figure 6. The sound source at the glottis does not have a cross section equal to that of the tract, however. So Figure 6 shows the impedance of the open cylinder, with the higher value of α, as would be measured through a glottal aperture of radius 3.2 mm (Hoppe *et al*. 2003), including its flanged end effect. Larger radii are also shown to illustrate the effect of widening the glottis. The first, low frequency minimum and maximum are not shown here because, viewed from the glottis end, the air in the tract cannot be treated as a compact object. The contribution of finite tissue compliance to the transfer functions with the open mouth will be reported in a later study. Note that no supra-glottal constriction is included here: such a constriction can have a considerable effect on the impedance at the glottis end of the tract, particularly on the higher formants.

The bandwidths of the models used plotted in the upper part of Figure 6 are shown in the lower part of the figure. The measurements in Figure 3 show that phonation further increases the bandwidths, perhaps because DC flow during phonation increases viscothermal losses, and possibly because of turbulence in the jet emanating from the glottis.

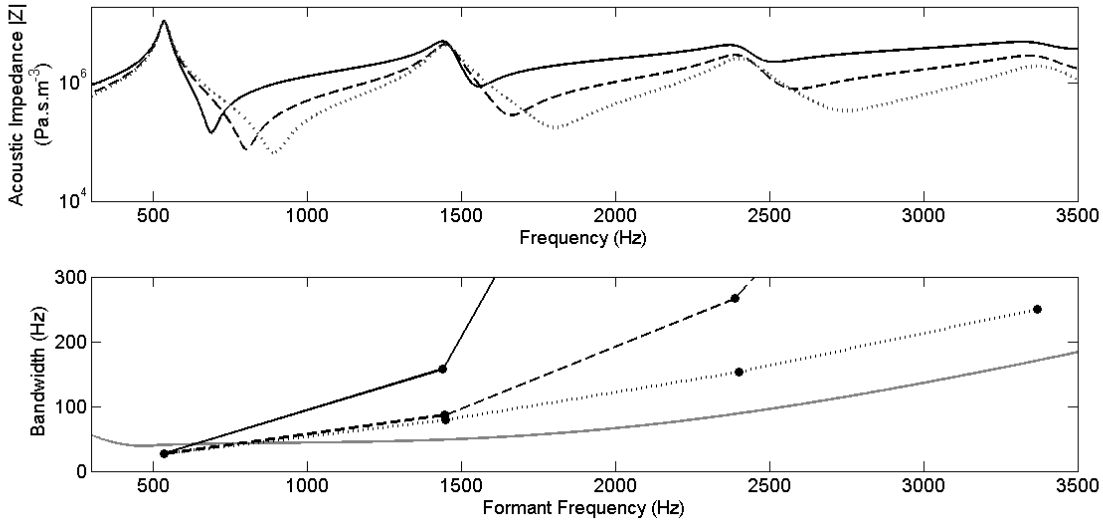


Figure 6. Calculated impedance spectra of a simplified vocal tract as seen at the open glottis, using the loss factor α estimated from Figure 4. The glottal aperture is modelled as the end effect of a tube of radius of 3.2 mm (solid), and 9.6 mm (dashed), and with the same radius as the vocal tract (dotted). The bandwidths of the maxima of these three models are shown in the lower plot. Note that with the smaller glottis openings some of the higher resonances have bandwidths that are undetermined. The smooth line shows an estimation of formant bandwidths for a phonation frequency of 150 Hz using the equation from Hawks and Miller (1995) (this curve would have large error bars not shown here).

The next step is to compare these measurements and calculations of bandwidth with the bandwidths of speech formants, the broad maxima in the spectral envelope of phonated speech. Formants cannot be precisely measured, because the frequency domain is only sampled at multiples of the fundamental frequency f_0. Hawks and Miller (1995) analysed previous measurements made with swept sine excitation and found a strong dependence on f_0, which may be related to this sampling. They published a formant bandwidth estimation curve, shown as an example in the lower part of Figure 6.

The preliminary results reported here will soon be augmented by a much larger data set and a more detailed analysis, which will be presented in the oral version of this paper. These will allow us to relate the bandwidths of formants to measured acoustical losses of the tract. The importance of such a comparison is considerable, because the value of the losses have important implications for perception of speech: very high losses would give broad resonances and broad but weak formants. Low losses would give narrow, strong resonances but, if the bandwidths were too narrow, few or no harmonics would be boosted by them. For example, at a steady phonation frequency of 150 Hz, a bandwidth of 50 Hz implies that there is a one-in-three chance that the harmonics of the voice will experience a substantial amplitude boost from a resonance of the vocal tract. In speech, where the fundamental frequency of the voice often changes during the course of a phoneme, the chances of experiencing a boost are somewhat greater.

The low frequency measurements provide data relating to the mechanical properties of the tissue and its interaction with the tract acoustics. The lowest mechanical resonance at ~20 Hz has a very low Q factor ~1, suggesting that the vibration of the tissue acts as a mass on a very lossy spring. The resonance maximum at ~200 Hz has a Q factor that is still fairly low: ~2. Although the 'spring' involved is that of the air, the motion of the tissue is also damped by the losses in deforming the tissue itself.

CONCLUSIONS

The frequency, magnitude and bandwidths of vocal tract resonances were measured directly both with the glottis closed and during phonation.

For an articulation approximating the vowel [ə:], the impedance magnitudes of purely acoustic resonances were 20-100 kPa.s.m^{-3} and antiresonances 2-10 MPa.s.m^{-3}. The bandwidths of these resonances with closed glottis and in the absence of radiation loss from the mouth were approximately 50 Hz. They depend only weakly on the resonance frequency.

The simple 1-dimensional vocal tract model approximates the measured data above 250 Hz if the losses are set to be several times higher than the visco-thermal wall losses expected for a rigid tube.

At lower frequencies, the vocal tract walls are not rigid and affect strongly the acoustics of the tract. A low Q factor impedance maximum due to the mechanical resonance of the tissue walls of the vocal tract was observed at ~200 Hz. An impedance minimum at ~20 Hz (mouth closed) is due to a resonance of tissue mass and the compliance of the air inside the tract.

ACKNOWLEDGEMENTS

We thank our volunteer subjects and the Australian Research Council for their support. The presentation of this paper was supported by a travel award from the New South Wales division of the Australian Acoustical Society.

REFERENCES

Brown, NJ, Xuan, W, Salome, CM, Berend, N, Hunter, ML, Musk, AW, James, AL & King, GG 2010, 'Reference equations for respiratory system resistance and reactance in adults', *Respiratory physiology & neurobiology*, 172(3): 162-168.

Childers, DG & J. A. Diaz 2000, 'Speech processing and synthesis toolboxes', *Journal of the Acoustical Society of America*, 108: 1975.

Childers, DG & Wu, K 1991, 'Gender recognition from speech. Part II: Fine analysis', *Journal of the Acoustical Society of America*, 90(4): 1841-1856.

Dickens, P, Smith, JR, & Wolfe, J 2007, 'Improved precision in measurements of acoustic impedance spectra using resonance-free calibration loads and controlled error distribution', *Journal of the Acoustical Society of America*, 121(3): 1471-1481.

Dunn, H 1961, 'Methods of measuring vowel formant bandwidths', *Journal of the Acoustical Society of America*, 33: 1737-1746.

Epps, J, Smith JR & Wolfe, J 1997, 'A novel instrument to measure acoustic resonances of the vocal tract during phonation', *Measurement Science & Technology*, 8(10): 1112-1121.

Fant, G 1970, *Acoustic theory of speech production with calculations based on X-ray studies of Russian articulations* Mouton De Gruyter.

Fant, G 1972, 'Vocal tract wall effects, losses, and resonance bandwidths', *Speech Transmission Laboratory Quarterly progress and status report*, 2(3): 28-52.

Fant, G, Nord, L & Branderud, P 1976, 'A note on the vocal tract wall impedance', *Speech Transmission Laboratory Quarterly progress and status report*, 4(1976): 13-20.

Flanagan, JL 1972, *Speech analysis; synthesis and perception* Berlin, New York, Springer-Verlag.

Fujimura, O & Lindqvist, J 1971, 'Sweep-Tone Measurements of Vocal-Tract Characteristics', *Journal of the Acoustical Society of America*, 49: 54 -558.

Hawks, JW & Miller, JD 1995, 'A formant bandwidth estimation procedure for vowel synthesis', *Journal of the Acoustical Society of America*, 97(2): 1343-1444.

Henrich, N, Smith, JR & Wolfe, J 2011, 'Vocal tract resonances in singing: Strategies used by sopranos, altos, tenors, and baritones', *Journal of the Acoustical Society of America*, 129: 1024-1035.

Hoppe, U, Rosanowski, F, Döllinger, M, Lohscheller, J, Schuster, M & Eysholdt, U 2003, 'Glissando: laryngeal motorics and acoustics', *Journal of Voice*, 17(3): 370-376.

Joliveau, E, Smith, J & Wolfe, J 2004a, 'Tuning of vocal tract resonance by sopranos', *Nature*, 427; 116.

Joliveau, E, Smith, J & Wolfe, J 2004b, 'Vocal tract resonances in singing: The soprano voice', *Journal of the Acoustical Society of America*, 116(4): 2434-2439.

Peterson, GE & Barney, HL 1952, 'Control methods used in a study of the vowels', *Journal of the Acoustical Society of America*, 24(2): 175-184.

Pham Thi Ngoc, Y & Badin, P 1994, 'Vocal tract acoustic transfer function measurements : further developments and applications', *J. Phys. IV*, C5: 549–552.

Rabiner, LR & Schafer, RW 1978, *Digital processing of speech signals* Prentice-hall Englewood Cliffs, NJ.

Rothenberg, M 1981, 'An interactive model for the voice source', *Speech Transmission Laboratory Quarterly progress and status report*: 1-17.

Savitzky, A & Golay MJE 1964, 'Smoothing and differentiation of data by simplified least squares procedures', *Analytical chemistry*, 36(8): 1627-1639.

Smith, JR 1995, 'Phasing of Harmonic Components to Optimize Measured Signal-to-Noise Ratios of Transfer-Functions', *Measurement Science & Technology*, 6(9): 1343-1348.

Sondhi, MM 1974, 'Model for wave propagation in a lossy vocal tract', *Journal of the Acoustical Society of America*, 55: 1070-1075.

Sondhi, MM 1986, 'Resonances of a bent vocal tract', *Journal of the Acoustical Society of America*, 79(4): 1113-1116.

Summerfield, Q, Foster, J, Tyler, R & Bailey, PJ 1985, 'Influences of formant bandwidth and auditory frequency selectivity on identification of place of articulation in stop consonants', *Speech Communication*, 4(1-3): 213-229.

Van den Berg, J 1955, 'Transmission of the vocal cavities', *Journal of the Acoustical Society of America*, 27: 161-168.

Wolfe, J, Garnier, M & Smith, J 2009, 'Vocal tract resonances in speech, singing, and playing musical instruments', *HFSP Journal*, 3(1): 6-23.

THE PLAYER–WIND INSTRUMENT INTERACTION

Joe Wolfe **André Almeida** **Jer Ming Chen** **David George** **Noel Hanna** **John Smith**
School of Physics, The University of New South Wales, Sydney
J.Wolfe@unsw.edu.au andregoios@gmail.com jerming@phys.unsw.edu.au
david.george@student.unsw.edu.au n.hanna@unsw.edu.au john.smith@unsw.edu.au

ABSTRACT

Players control a range of parameters in the player-instrument system. First we show how loudness and pitch vary over the plane of mouth pressure and force on the reed of a clarinet, and thus how these parameters can be used in compensation to produce trajectories in this plane that have varying loudness and timbre but constant pitch. Next we present impedance spectra for several different types of musical instruments and for the vocal tract, to allow general observations. We report different ways in which the acoustic properties of the player's tract interact with those of the instrument bore to control the frequency of reed vibration in some wind instruments. We also show how vocal tract resonances can influence timbre.

1. INTRODUCTION

The player often is more interesting than the instrument: in general, a good musician on a poor instrument sounds better than a poor musician on a good instrument. Some of what makes a musician good lies outside the realm of science, let alone acoustics, but acousticians can contribute to understanding good performance by researching the player-instrument interaction, aiming to understand how good players achieve their musical goals. Beyond its intrinsic interest, this has possible applications: an improved and explicit understanding of how good players play could help guide students and teachers.

Some of the player-instrument interaction consists in the former doing something to the latter: pushing the right keys in the right sequence, applying certain values of force with the lips, certain pressures from the lungs, etc. These are of continuing interest in our lab, and we begin by presenting one such set of parameters.

Another part of the player-instrument interaction for wind instruments is directly acoustic in that it involves the acoustics of the player's vocal tract.

1.1 The elements

Performance involves the interaction of the principal acoustical elements of the wind instrument–player system: (i) a source of air at positive pressure, (ii) a vibrating element, usually an air jet, a reed or the player's lips, (iii) the downstream duct, *i.e.* the bore of the instrument and (iv) the upstream duct, comprising the player's airway.

Players control all of these: (i) The air pressure and flow are controlled by muscles of the torso and also, in some cases, by the glottis. On the very short time-scale, flow is also controlled in an almost binary fashion by the tongue, which can cease the flow by contact with the roof of the mouth (in gestures like *ta*, *la*, *da*, *ka* etc) or by contact with the reed. (ii) The valves are diverse: Flutists control lip aperture size and geometry. Brass players vary the geometry and mechanical properties of the lips. Reed players choose or make their reeds and vary their mechanical properties with their lips. (iii) The geometry of the downstream duct is varied with valves and slides in brass instruments and by opening or closing covering tone holes in woodwind. (iv) In many cases, players vary the shape and position of the tongue, palate and the opening of the glottis, to control the acoustic properties of the upstream tract.

Two teams discussed the acoustics of the upstream duct at the last SMAC [1,2] and a friendly rivalry began. Scavone [1] presented circuit models of an upstream resonator and a downstream waveguide. We reported results using physical models of the vocal tract and either a cantilever valve or water filled latex 'lips' to represent player's lips [2]. We next used a broad-band signal and the capillary technique to measure the acoustic impedance in the mouth of a player while he played the didjeridu [3]. Later, we applied the technique to saxophone, clarinet and trumpet [4-6]. Meanwhile, Scavone developed a different technique to study the tract involvement: microphones inside bore and mouth give the ratio of the two impedances for harmonics of the note [7]. The two techniques are somewhat complementary and the teams progressed in parallel: Scavone's technique uses the vibrating reed as the (large) signal, which gives large signal:noise ratio and so allows rapid measurements and the ability to track rapid changes in time. However, because it only samples the frequency domain at multiples of the playing frequency, involvement of the vocal tract support must be inferred from a sparse representation (it does not directly measure tract resonances). Our technique gives the impedance spectrum in the mouth but, because our probe signal's energy is spread over hundreds of frequencies, we need windows of tens of ms up to seconds.

In this paper, we review aspects of the musician's control in all four areas and present new work. The review disproportionally cites our own work, in part because the player-instrument interaction has been one of the main lines of research in our lab since the last SMAC.

2. COMPENSATING CONTROL PARAMETERS

In many instruments, increasing pressure in the mouth, with all other parameters held constant, changes the pitch. On a recorder, the only parameter that the player can use to control the jet is the mouth pressure. Consequently, it is difficult to play a *decrescendo* on a sustained note without a fall in pitch. Some accomplished players compensate for these pitch changes using the downstream duct by partially opening and closing tone holes.

In other instruments, the effect of changing pressure may be compensated with other parameters to keep the pitch constant. In reed instruments, for a given configuration of the bore, the pitch is also dependent on the forces applied to the reed and to the configuration of the upstream duct. Consequently, a player can play a *decrescendo* at constant pitch by following an iso-frequency contour through the space of control parameters.

The clarinet is the 'lab rat' of wind instruments: many scientific studies and models exist [*e.g.* 8-10]. One way to study control parameters is to measure some of them in the player as in [11] or in a blowing machine, as in [12]. Here we use a systematic study of on how pitch, loudness and spectrum depend on mouth pressure, lip force and reed parameters. To control these parameters and to maintain a constant configuration of the upstream duct, we use an automated clarinet playing system.

2.1 The automated clarinet player

The clarinet player was built partly for this purpose [13] and partly for a competition (Artemis International) for such automata. A film clip of the player in concert is at www.phys.unsw.edu.au/jw/clarinetrobot.html. A pump provides air whose pressure is regulated in the long term by controlling the pump and on short time scales via a controlled leak. In this experiment, the reed force was applied by the weight of suitable masses through a soft sorbothane pad, which provided losses somewhat like those from a human lower lip. The reeds are synthetic (Légère, Barrie, Ontario) with 'hardness' ratings 2 and 3.5, the clarinet a Yamaha YCL250 with a Yamaha CL-4C mouthpiece.

2.2 Pitch and loudness control

Figure 1 plots the frequency f and sound level L as functions of the gauge pressure P in the 'mouth' and the force F applied by the 'lip'. The force is applied at 10 mm from the reed tip. The shaded region represents the parameter combinations that produce notes, except that the upper pressure limit near 7 kPa is imposed by the pump used. The reed has a manufacturer's 'hardness' rating of 2, the note played is F4 (written G4) and the frequencies are a little low because the instrument is playing at 19±1 °C. This note is in the clarinet's 'throat register': it is an octave and a minor third above the lowest note, it is still in the first mode of vibration, and 12 tone holes are open.

Above the shaded region, no sound is produced because the combination of lip force and pressure push the reed closed against the mouthpiece. This extinction line has a negative slope with a magnitude of a few cm^2. Left of the edge of the playing region, with a rather steeper slope, either no sound or squeaks are produced. At very low F, the damping is insufficient to prevent squeaks: *i.e.* sounds with frequencies close to the resonance of the reed, rather than the resonance of the bore.

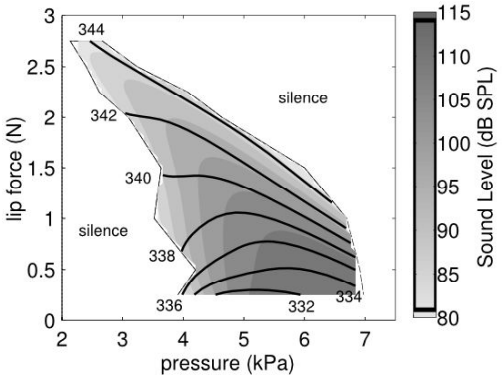


Figure 1 In the space of pressure in the artificial mouth and force exerted on the reed, we plot frequency in Hz (black lines) and sound pressure level (shading) measured at 50 mm from the bell on the axis of the clarinet.

f is higher at higher F and also, over most of the playing regime, at higher P. The iso-f lines have negative slopes of a few cm^2 over most of the playing regime. This is qualitatively explained by a simple effect: Both F and P tend to close the reed against the mouthpiece, which reduces the compliance of the air in the mouthpiece and also the effective compliance of the reed. This raises the frequency of the peaks in the parallel impedance of reed and bore.

Figure 1 shows that, over part of the range, a player could vary sound level and constant frequency by blowing harder and making a compensating reduction in lip force, and adjusting no other parameters. At high P and F, however, the iso-f lines and the iso-L contours are nearly parallel. So, if no other control parameters were used, the player would need to fall to a lower pitch (by relaxing the jaw (lower F)) to play more loudly. Going flat when playing loudly is a fault often identified by clarinet teachers.

In most instruments, vibrations of increased amplitude produce greater nonlinearities in the valve, with the consequence that the amplitudes of higher harmonics increase more rapidly than those of the fundamental. A simple quantification of this is the spectral centroid, the frequency weighted average of the amplitude of all spectral components. The spectral centroid is strongly correlated with the perceived brightness of timbre. Contours of the spectral centroid on the P,F plane have shapes similar to those of sound level (data not shown).

F4, the note shown in Figure 1 is one of the easiest to sound on the clarinet. For notes one octave below and one octave above, the playing regime is rather smaller. Lip force application at 10 mm from the reed tip facilitates sound production: in experiments with the lip force applied 5 mm either side of this position, the playing regimes are smaller. Finally, in similar experiments using a stiffer reed, the results are qualitatively similar, but notes are only produced in the higher range of F, above about

3 N. Space precludes including these data here: they have been submitted for publication elsewhere [14].

3. THE UPSTREAM AND DOWNSTREAM DUCTS

In some cases, there is a limited symmetry between the two ducts [15]. Call the total and acoustic pressures immediately upstream from the reed, inside the player's vocal tract, P_{mouth} and p_{mouth} and use the subscript bore for those immediately downstream, in the duct provided by the instrument. A clarinet reed tends to bend inwards towards the mouthpiece, thus tending to close the flow pathway, when the pressure difference $P_{\text{mouth}} - P_{\text{bore}}$ is positive. Under some playing conditions, $P_{\text{mouth}} - P_{\text{bore}}$ acts to bend a brass player's lips outwards into the instrument, opening the flow pathway [16,17]. Defining the acoustic impedances Z_{mouth} and Z_{bore} as p/U, where U the acoustic volume flow into the duct then, if the flows out of the mouth ($-U_{\text{mouth}}$) and into the instrument (U_{bore}) are equal, say U, then $p_{\text{mouth}} - p_{\text{bore}} = -UZ_{\text{mouth}} - UZ_{\text{bore}} = -U(Z_{\text{mouth}} + Z_{\text{bore}})$. In words, Z_{mouth} and Z_{bore} are in series with regard to the mechanisms described. It is therefore worth comparing and contrasting the impedance spectra of some of the ducts involved, which we do in the next section.

It can also be shown that the passive impedance of a reed or lip is in parallel with the series combination mentioned above. Further, the pressure difference $P_{\text{mouth}} - P_{\text{bore}}$ across the valve is not the only source of force acting to open or close it: for instance, the dynamic or 'Bernoulli' pressure $\frac{1}{2}\rho v^2$ can also play a role, so the two ducts are not necessarily in series with respect to all possible regeneration mechanisms in the valve.

3.1 Impedance measurements

With one exception, the impedance spectra shown in Figure 2 are made using the three-microphone technique calibrated with three non-resonant loads, one of which is an acoustically infinite duct [18]. The smallest microphone spacing in this impedance head is 40 mm, which limits the frequency response to about 4 kHz. The lower limit is about 10 Hz, giving a range of 9 octaves. The use of non-resonant calibration loads and the iterative optimisation of the measurement signal together allow a dynamic range of more than 90 dB. The frequency resolution depends on the period of the measurement signal.

3.2 Ducts, resonances and antiresonances

Figures 2, 3 and 4 show the impedance spectra Z of some simple ducts, several instruments and a vocal tract, to allow some general discussion. Figure 2 shows the measured impedance spectra of a number of ducts. (a) shows the impedance of an open cylinder with an internal diameter of 15 mm. (b) shows an open cone with a half angle of 1.74° (equal to that of a soprano saxophone) [19]. The cone is truncated at the small (input) end to allow flow into the cone via the impedance head used to measure it, and the truncation replaced with a cylinder of equal volume. Their effective lengths are 325 mm, chosen so that the first maximum in Z of the cylinder occurs at C4 and the first minimum at C5.

3.3 The downstream duct: woodwinds

The player's control of the downstream duct often involves complicated co-ordinations, such as manipulating different keys in the same transition. Departures from simultaneity are sometimes systematic [20], which raises interesting acoustic and pedagogical questions, which we shall not pursue further here, but instead concentrate on quiescent states.

Below the cylinder in Figure 2 we show Z for a flute [21] and clarinet [22], whose bores include cylindrical sections. (To save space, phase is not shown for these curves.)

The array of open tone holes creates a cut-off frequency, which is about 1.5 kHz for the clarinet. Below this frequency, sound waves are reflected at a point near the first of the open holes. Well above this, they travel down the whole length of the bore, which explains why the spacing of the peaks in Z decreases above that frequency. That cut-off for the flute is around 2 kHz but the more closely spaced peaks are less evident, because at several kHz the bore is short circuited by a shunt provided by the Helmholtz resonator in the head joint [21].

The observation that there is only a little similarity between Z for these instruments and that for a cylinder explains why highly simplified models fail: for example, the statement that clarinets have weak even harmonics is not usually true.

Below the cone are shown Z for soprano and tenor saxophones [19]. The soprano sax, the clarinet, the flute and the two simple geometries all have the same effective length, so all play C5 in this configuration (dashed circles) except the clarinet, which plays C4.

The tenor saxophone is shown on an expanded scale to show that it is roughly a 1:2 scale model, playing an octave below, although the cone half-angles are different: 1.74° and 1.52° respectively.

For the saxophones and clarinet, the dotted line shows the bore impedance in parallel with the compliance of a typical reed. This is important in intonation: a soft reed (as much as twice the compliance used here) lowers the peak of the parallel impedance and the instrument would play flat unless the tuning slide were shortened.

Note that the saxophones have only two large peaks in Z. The clarinet has three and, above the cut-off frequency, further sharp peaks, whose narrower frequency spacing indicates that they are standing waves over the whole length of the bore. The conical bores of saxophones give little reflection for high frequencies, and so only weak standing waves. As we'll see later, this makes the vocal tract especially important for the high range of the saxophone.

3.4 The downstream duct: lip valves

The Z for four lip valve instruments are shown in Figure 3. (To conserve space, phase is not displayed.)

The first is a didjeridu. The first several extrema are qualitatively consistent with a slightly flared duct, as the

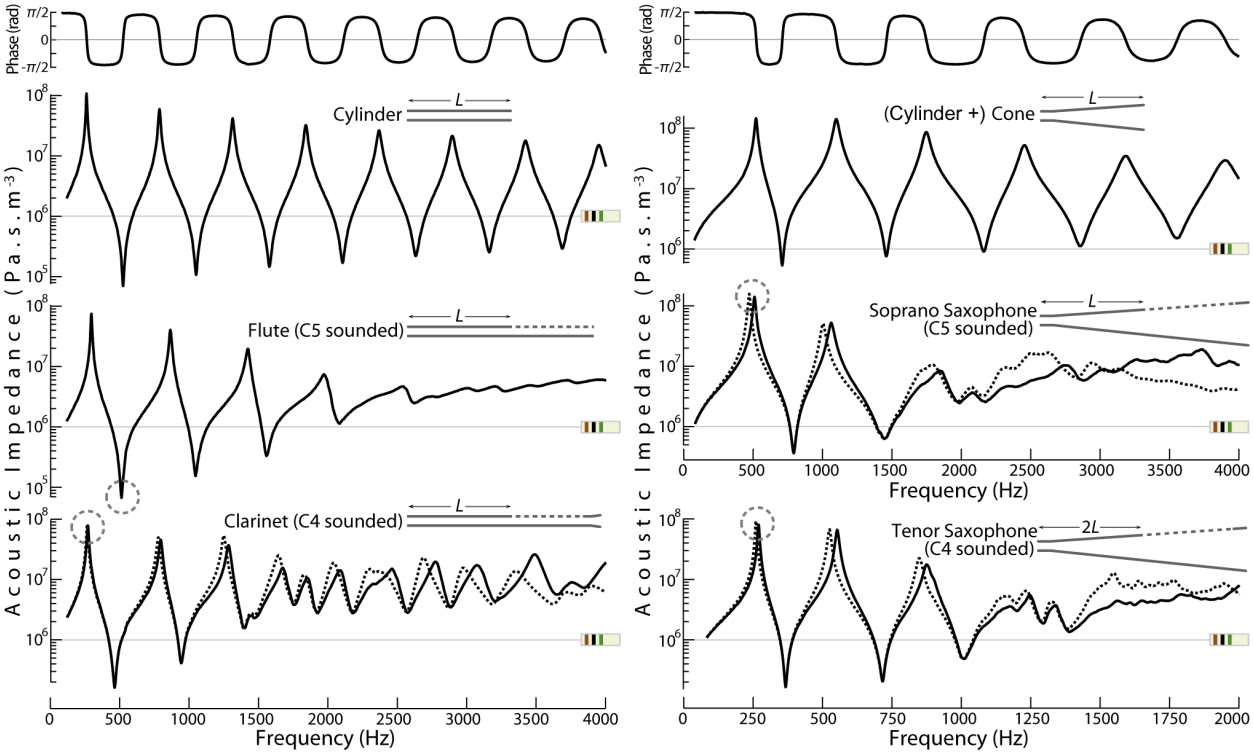


Figure 2. Impedance spectra. At top, a cylinder (left) and a truncated cone (right). These have the same effective length as the flute and soprano saxophone (fingering for C5) and the clarinet (C4). The tenor saxophone has the same fingering as the soprano but is plotted on an expanded frequency axis to show that it is roughly a 1:2 scale model. (Some phase plots are omitted to save space.)

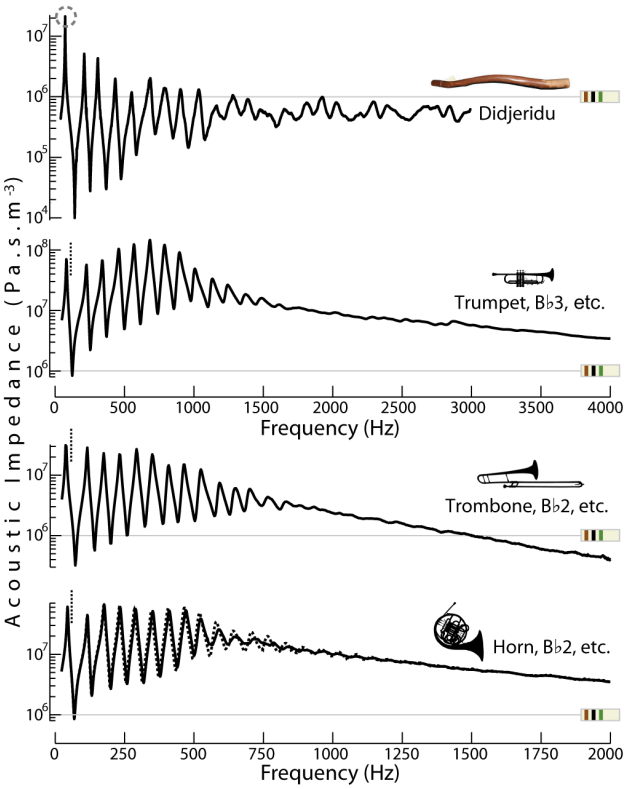


Figure 3. Impedance spectra of lip-valve instruments: a didjeridu, trumpet, trombone and horn: the last two on expanded scales as they are approximately 2:1 scale models of the trumpet. On Figures 2, 3 and 4, the 1 MΩ line is marked.

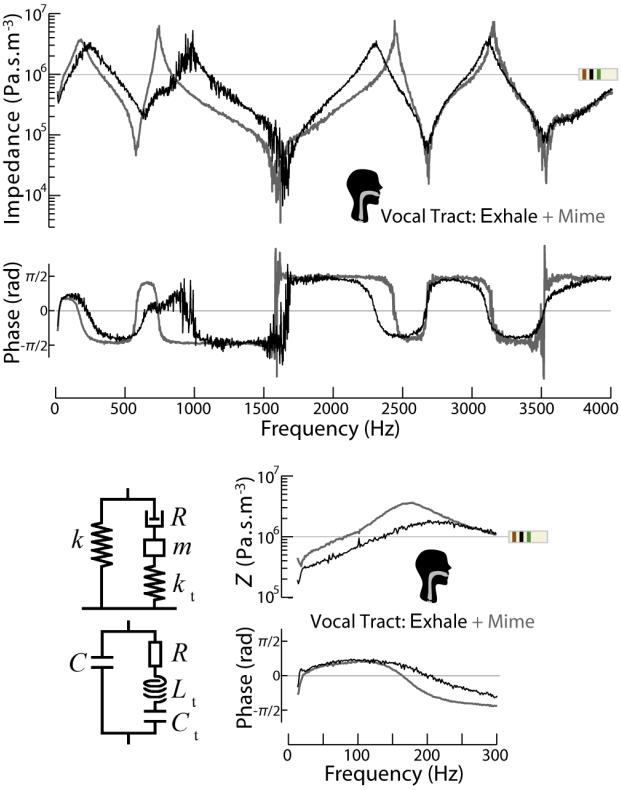


Figure 4. Impedance spectra of a vocal tract measured at the lips over 370 ms: glottis closed (pale) and exhaling (black). The compact-object equivalent circuit is shown next to the plots of low frequency behaviour on an expanded axis.

outer shape suggests. The impedance peak at which it plays is indicated (dashed circle). The irregular features in Z and the absence of strong resonances at high frequencies are due to the irregular interior, which is made by termites in a branch or trunk of a eucalypt tree.

The Z for trombone (slide in) and horn (B♭ side, no valves depressed) are shown on an expanded scale to show that they are, very approximately, scale models of the trumpet (no valves depressed). In all three cases, the cut-off frequency is provided by radiation from the bell, which becomes important when the wavelength is comparable with the radius of curvature of the bell. For the horn, the hand in the bell (dotted line) reduces the efficiency of radiation and so raises the cut-off frequency and the height of the higher frequency peaks.

The three brass instruments each have about 10 or 12 impedance peaks. Starting from the second, these are approximately equally spaced. In this configuration, notes near all of those peaks are played. They correspond to a harmonic series whose fundamental is half the frequency of the second peak. A note with that pitch can be played as a pedal note (dashed vertical line). (The first peak, well below the pedal pitch, is not played.)

Comparing these plots, we might expect that considerably more control over the valve is required of a brass player, whose impedance peaks are more closely spaced. In contrast, when the saxophonist and clarinetist wish to use the second (or third), weaker peak, they can simply use register key(s) to weaken and to detune the lower peak(s).

3.5 The upstream duct: the vocal tract

Figure 4 shows Z for a vocal tract. The impedance head has a diameter of 26 mm and the lips are sealed around that, as described in [23]. The tongue is in the position to pronounce /əː/ as in 'heard'. For the grey curve, the glottis is closed, so there is no DC flow. For the black curve, the subject is exhaling into the impedance head, which has a downstream vent for this purpose.

Nine octaves covers the first five resonances. Because of the long period of the first resonance, the measurements were made over a single window of 370 ms (= 2^{14}/44.1 kHz). For the exhalation case, this shows a measurement limitation: noise due to the turbulent air flow is superposed over the broad-band measured signal. Integration over longer time-scales can improve this, but the subject must sustain the gesture for longer.

Opening the glottis (going from a closed to a somewhat open pipe), one expects extrema in impedance to rise slightly in frequency. For example, whisper uses a larger glottal opening than normal speech, and measurements of the tract resonances from normal phonation to whisper in the same gesture show an increase in frequency [24]. The measurements in Figure 4 were not measured in the same gesture, so other geometrical changes may also contribute, particularly at high frequencies. It is believed that players use a slightly open glottis [25], so a playing configuration could be between these two conditions.

If the vocal tract were a rigid cylinder, 170 mm long and closed at the glottis, we should expect minima at 0.5, 1.5, 2.5 and 3.5 kHz, and maxima at about 1, 2 and 3 kHz, which is roughly what we observe.

In a closed, rigid cylinder, Z would be very large at very low frequencies. Human tissues are not infinitely rigid, of course, and this gives rise to the zeroth minimum and maximum, shown in the inset in Figure 4. The maximum at about 200 Hz is due to the mass of the tissue surrounding the tract and the 'spring' of the air inside it, which is approximately a compact object at this wavelength (2 m).

The minimum at about 20 Hz is due to the mass of the tissue and the 'spring' of its own elasticity. Because the two resonances are three octaves or more apart, impedance maxima and minima occur at $\omega \sim 1/(LC)^{½}$ and $1/(LC_t)^{½}$ respectively, where L is the inertance of the tissue, C_t the compliance due to the supporting tissues, and C the compliance of the air in the upper tract. Taking a volume of 100 ml for the air in the tract and a surface area of the surrounding tissues of 10^{-4} m^2, L gives a tissue thickness of ~1 cm and C_t a spring constant of ~1/N.cm^{-1}.

3.6 Varying the resonances of the upstream duct

How much can the upstream resonances be adjusted by articulation? Opening and closing the glottis makes little difference at high frequency, because the inertance of air in the glottis effectively seals it. Of course, other geometrical changes associated with this variation also may make a difference. In speech, the first resonance is varied primarily by varying the opening at the lips, which connects the tract to the very low impedance of the radiation field. This option is not available to reed and brass players, whose mouths make an airtight seal to the instrument. So the first resonance cannot be varied much and, in our measurements on a range of instruments, we usually see an acoustic maximum between about 200 and 400 Hz, and a minimum a few hundred Hz above that.

The shape of the vocal tract can be used to adjust the resonant frequencies. This is most effective once the half wavelength becomes comparable with the length of the upper vocal tract – when the frequency approaches 1 kHz, but even the first maximum can be varied. As a general rule: a constriction near a pressure node lowers the frequency while one near a pressure antinode raises it.

Let's compare the magnitudes of impedance peaks in figures 3, 4 and 5: those in the vocal tract, with jaw low and tongue neutral, are several MΩ. This already approaches the values of the instrument impedance at high frequency. Further, players can raise this value by raising the tongue near the instrument, which makes the front of the mouth act like an impedance matching cone and produces rather higher impedances, as we shall see. It is easier for an impedance peak in the tract to compete with one in the bore at high frequencies, where the bore resonances are weaker.

4. VOCAL TRACT EFFECTS

4.1 Tract-bore series combination can control pitch

A reed and the air that flows past it are driven by the pressure difference between mouth and bore, which

means that, as described above, the tract and bore act in series.

For the saxophone, this solves a problem: Figure 2 shows that, because of effective radiation and weaker reflection from the large end of the cone, the instrument's Z has only two strong impedance peaks. Although the first register is extended downwards in pitch with the use of extra keys, and the second similarly upwards, the range of the instrument using these first two peaks is just 2 octaves and a musical fifth.

Figure 5 contrasts the situation for a note, G4, in the second register of a tenor saxophone and A#5, in the *altissimo* range. Even beginners can play the first note, with relatively little attention to control parameters. The *altissimo* notes require expertise, because they are only possible when a large amplitude peak in Z_{mouth} is produced and tuned to select one of the instrument's weaker high resonances [4].

Advanced saxophonists can also use large magnitude peaks in Z_{mouth} to select and to adjust the amplitude of notes in mulitphonics (chords) and also to 'bend' the pitch of notes [26].

Because its bore is partly cylindrical, the higher resonances in the clarinet produce stronger peaks in Z_{bore} (Figure 2), so less assistance from the upstream duct is required for altissimo playing. However, like saxophonists, clarinetists use peaks in Z_{mouth} in multiphonics and in pitch bending, including the famous *glissando* in *Rhapsody in Blue* [5].

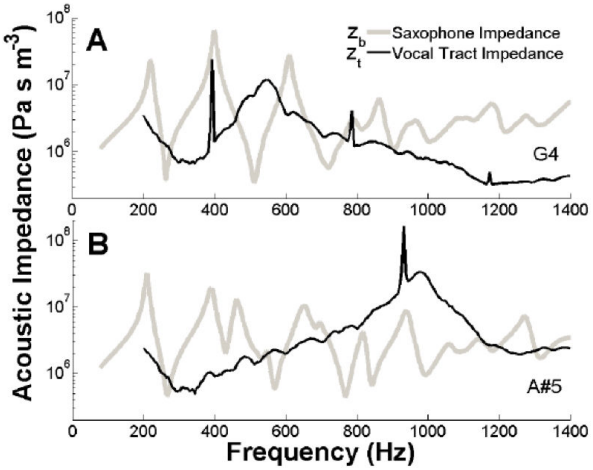


Figure 5. The bore (grey) and vocal tract impedance (black) for two notes played on a tenor saxophone. The narrow peaks are harmonics of the notes played. G4 is in the instrument's second register. A#5 is in the *altissimo*, and is accessible only by tuning a peak in Z_{mouth} to select one of the weak, higher resonances. Reproduced from [4].

Figure 6 shows the vocal tract resonances of saxophonists, clarinetists and trumpeters. In the normal range of the saxophone, neither advanced nor less advanced players show tuning of peaks in Z_{mouth} to the played note. In the *altissimo* range, however, peaks are tuned either to or slightly above the note to be played and those unable to do this cannot play in this register.

For clarinetists, peaks in Z_{mouth} are tuned to the note to be played during pitch bending. In normal playing, however, peaks in Z_{mouth} are tuned a couple of hundred Hz above the note to be played, which means that the upstream impedance is inertive [5].

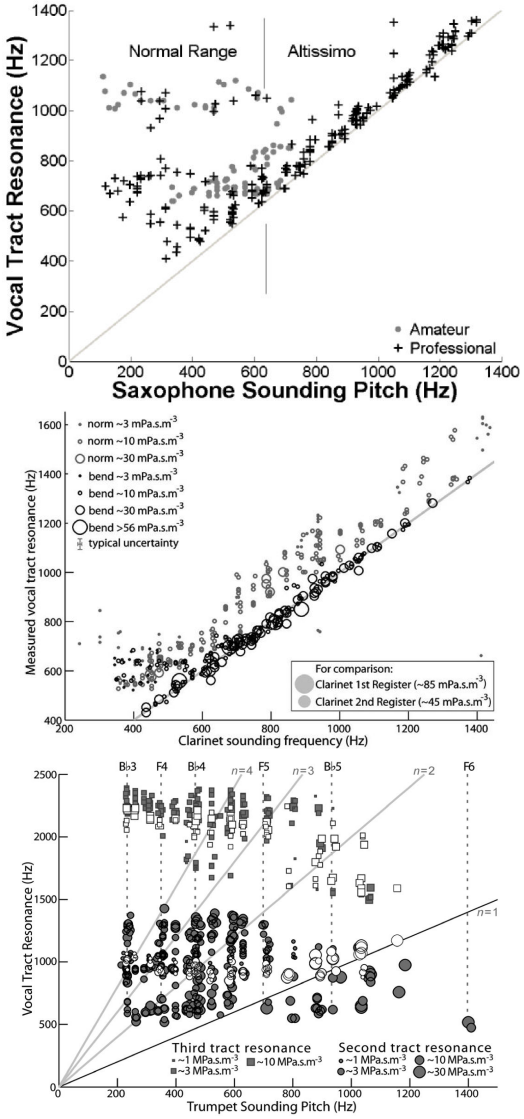


Figure 6. Vocal tract resonance frequency against pitch frequency. Top: tenor saxophone in normal (left) and altisimo ranges, showing resonance tuning in the latter. Middle: clarinet pitch bending (black) and normal playing (gray): resonance tuning for pitch bending, and the resonance held somewhat above the pitch in normal playing. Bottom: the seven trumpet players in this study show no consistent resonance tuning. Reproduced from [4,5,6].

What about trumpet players? Some players can play in the upper part of the third octave, where the peaks in peaks in Z_{bore} are weak. Do they tune a peak in peaks in Z_{mouth} to assist playing in this *altissimo* range? When we started this study [6], we expected to find vocal tract tuning like that of saxophonists. But the answer, in general, is no. Figure 6 shows no consistent tuning by the seven players in our study.

So, what are the effects of Z_{mouth} in players of brass and other lip-valve instruments? At the previous SMAC, we presented Figure 7, the results of a trombone-playing system using, as the 'lip-valve', an outward swinging cantilever spring with a mass added to give a suitable natural frequency. Upstream, we used different shaped simple cavities as models of the mouth with tongue low in the mouth and high in the mouth. The trombone slide was then moved into its standard positions for notes on the chromatic scale. While the valve operated on the same peak in Z_{bore} (i.e. played in the same register), the pitch decreased by approximately a semitone for each step increase in the slide position, as expected. Depending on the natural frequency of the valve, the system would jump from one register to the next at a particular position.

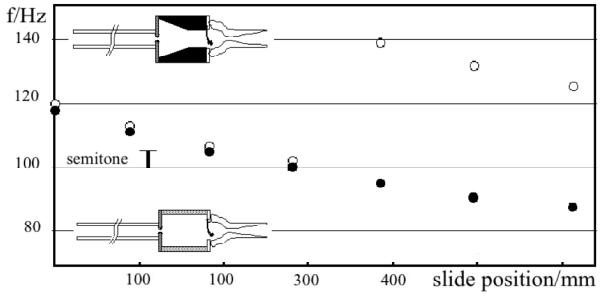


Figure 7. A trombone is 'played' with different slide positions by an artificial valve with two upstream cavities with different shapes.

Without the constriction in place ('low tongue'), the system played the same register flatter than with the constriction in place ('high tongue'). Further, the jump between registers occurred at lower pitches without the constriction. This is consistent with players' observations that lowering the tongue either lowers the pitch slightly or else causes the instrument to jump to a lower register. More recent measurements of the player-trombone interaction are given in another paper in this volume [27].

4.2 Vocal tract interactions with higher harmonics

In the saxophone and clarinet, impedance peaks in Z_{mouth} can contribute to reed vibration, either on their own or in collaboration with peaks in Z_{bore}. Peaks in Z_{mouth} can also, of course, inhibit acoustic flow. Which of these applies?

Players of various wind instruments report that they use different configurations of the vocal tract to control timbre. This is most spectacularly evident on the didjeridu, an indigenous Australian instrument, which is played almost entirely at its lowest resonance. Its musical interest comes from rhythmic variations in timbre produced, *inter alia*, by varying vocal tract shapes, including those used for the cyclic breathing that allows the didjeridu to be played continuously without pauses for breath.

At the last SMAC, we also presented this sound file www.phys.unsw.edu.au/jw/sounds/dij_trombone.wav. It was produced by a pair of artificial lips (water-filled latex cylinders) with a cylindrical pipe downstream modelling the didjeridu. Upstream was a model vocal tract with continuously variable resonances.

Figure 8 shows the spectrum of sound radiated from a cylindrical pipe being played as a didjeridu by a human player. While the subject played, a small impedance head placed between the lips measured Z_{mouth} using the capillary method. On the same figure, vertical tics indicate the harmonics of the note played and the resonances of the pipe. Unlike a typical clarinet note, the first eleven odd harmonics in this note really are all weaker than their even neighbours, though of course this is not the case with a real didjeridu. The spectral envelope shows minima around 1.5 and 2 kHz. The lower part of the figure shows that Z_{mouth} has maxima at these frequencies: at these frequencies, very little current can pass between the lips, so there is little power input to the instrument, and so little power in the radiated sound.

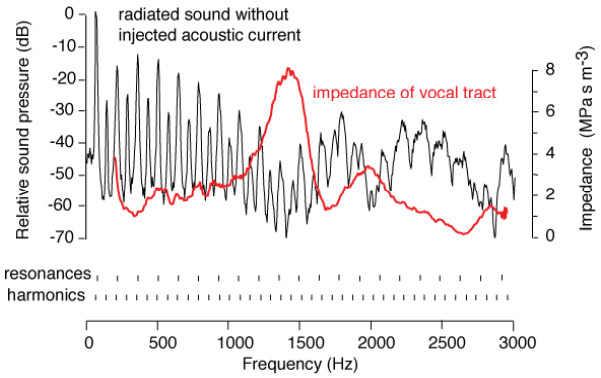


Figure 8. A cylindrical pipe played like a didjeridu. The spectrum of the radiated sound, the frequencies of the pipe resonances and the harmonics and the impedance spectrum of the vocal tract.

The didjeridu has no constriction comparable to that in the mouthpiece of a typical brass instrument and the magnitude of its Z_{bore} is, overall, rather less than those of typical woodwind or brass instruments (see Figures 2-4). Are similar effects observed on other wind instruments? In other papers in this volume, we report Z_{mouth} measurements for the saxophone and the trombone [27,28].

5. CONCLUSIONS

Players must simultaneously control several parameters simultaneously in order to produce desired contours of pitch, loudness and timbre. Features of the acoustic impedance of the vocal tract contribute to pitch, especially when an impedance peak lies near the playing frequency, and to timbre, when peaks fall close to higher harmonics.

Acknowledgments

We thank the Australian Research Council for support, Yamaha for instruments and our volunteer subjects.

6. REFERENCES

[1] G.P. Scavone, "Modeling vocal-tract influence in reed wind instruments," in Proc. Stockholm Music Acoustics Conf. (SMAC 03) 2003, Stockholm, pp. 291–294.

[2] J. Wolfe, A.Z. Tarnopolsky, N.H. Fletcher, L.C.L.

Hollenberg, and J. Smith, "Some effects of the player's vocal tract and tongue on wind instrument sound," in Proc. Stockholm Music Acoust. Conf. (SMAC 03), Stockholm, 2003, pp. 307-310.

[3] A. Tarnopolsky, N, Fletcher, L. Hollenberg, B. Lange, J. Smith, and J. Wolfe, "Vocal tract resonances and the sound of the Australian didjeridu (yidaki) I: Experiment", in J. Acoust. Soc. America, 2006, **119**, pp. 1194-1204.

[4] J.-M.,Chen, J. Smith, and J. Wolfe, "Experienced saxophonists learn to tune their vocal tracts," in Science, 2008, **319**, p726.

[5] J.-M.,Chen, J. Smith, and J. Wolfe, "Pitch bending and glissandi on the clarinet: roles of the vocal tract and partial tone hole closure," in J. Acoust. Soc. Am., 2009, **126**, pp. 1511–1520.

[6] J.-M.,Chen, J. Smith, and J. Wolfe, "Do trumpet players tune resonances of the vocal tract?" in J. Acoust. Soc. Am., 2012, **131**, pp. 722–727.

[7] G.P. Scavone, A. Lefebre, and A.R. Da Silva, "Measurement of vocal-tract influence during saxophone performance," in J. Acoust. Soc. Am., 2008, **123**, pp. 2391–2400.

[8] J. Backus, "Small-Vibration Theory of the Clarinet," in J. Acoust. Soc. Am., 1963, **35**, pp. 305-313.

[9] J. Gilbert, J. Kergomard, and E. Ngoya, "Calculation of the steady-state oscillations of a clarinet using the harmonic balance technique", in J. Acoust. Soc. Am., 1989, **86**, pp. 35-41.

[10] J.-P. Dalmont and C. Frappé, "Oscillation and extinction thresholds of the clarinet: Comparison of analytical results and experiments", in J. Acoust. Soc. America, 2007, **122**, pp. 1173–1179.

[11] Ph. Guillemain, Ch. Vergez, D. Ferrand and A. Farcy, "An instrumented saxophone mouthpiece and its use to understand how an experienced musician plays" in Acta Acust., 2010, **96**, pp.622-634.

[12] T.A.Wilson, and G.S. Beavers, "Operating modes of the clarinet", in J. Acoust. Soc. America, 1973, **56**, pp. 653-658.

[13] A. Alméida, J. Lemare, M. Sheahan, J. Judge, R. Auvray, K.S. Dang, S. John, J. Geoffoy, J. Katupitiya, P. Santus, A. Skougarevsky, J. Smith, and J. Wolfe, "Clarinet parameter cartography: automatic mapping of the sound produced as a function of blowing pressure and reed force" in ISMA2010, Sydney & Katoomba.

[14] A. Almeida, D. George, J. Smith, and J. Wolfe, "The clarinet: how blowing pressure, lip force and reed hardness affect pitch, loudness and timbre," in J. Acoust. Soc. America, submitted.

[15] A.H. Benade, "Chapter 35: Air column, reed, and player's windway interaction in musical instruments," in Vocal Fold Physiology, Biomechanics, Acoustics, and Phonatory Control, 1985, Denver, pp. 425–452.

[16] S.J. Elliott,, and J.M. Bowsher, "Regeneration in brass wind instruments," in J. Sound Vib., 1982, **83**, pp. 181–217.

[17] S. Yoshikawa, "Acoustical behavior of brass player's lips" in J. Acoust. Soc. Am., 1965, **97**, pp. 1929–1939.

[18] P. Dickens, J. Smith, and J. Wolfe, "High precision measurements of acoustic impedance spectra using resonance-free calibration loads and controlled error distribution," in J. Acoust. Soc. Am., 2007, **121**, pp. 1471–1481.

[19] J.-M.,Chen, J. Smith, and J. Wolfe, "Saxophone acoustics: introducing a compendium of impedance and sound spectra, " in Acoustics Australia, 2009, **37**, pp. 18-23.

[20] A. Almeida, R. Chow, J. Smith, and J. Wolfe, J. "The kinetics and acoustics of fingering and note transitions on the flute," in J. Acoust. Soc. Am, 2009, J. Acoust. Soc. Am., **126**, pp. 1521-1529.

[21] J. Wolfe, J. Smith, J. Tann, and N.H. Fletcher, "Acoustic impedance of classical and modern flutes" in J. Sound Vib., 2001, **243**, pp. 127-144.

[22] P. Dickens, R. France, J. Smith, and J. Wolfe, "Clarinet acoustics: introducing a compendium of impedance and sound spectra", in Acoustics Australia, 2007, **35**, pp. 17-24.

[23] N. Hanna, J. Smith, and J. Wolfe, "Low frequency response of the vocal tract: acoustic and mechanical resonances and their losses" in Proc. Acoustics 2012 Fremantle, Australia.

[24] Y. Swerdlin, J. Smith, and J. Wolfe, "The effect of whisper and creak vocal mechanisms on vocal tract resonances", in J. Acoust. Soc. America, 2010, **127**, pp. 2590-2598.

[25] M.S. Mukai, "Laryngeal movements while playing wind instruments," in Proc. Int. Symp. Music Acoust., 1992, Tokyo, pp. 239–242.

[26] J.-M.,Chen, J. Smith, and J. Wolfe, "Saxophonists tune vocal tract resonances in advanced performance techniques," in J. Acoust. Soc. Am., 2011, **129**, pp. 415–426.

[27] H. Boutin, N. Fletcher, J. Smith and J. Wolfe "The playing frequency of the trombone and the impedances of the upstream and downstream ducts" in SMAC13, 2013, Stockholm.

[28] Li, W.-C., J.-M. Chen, J. Smith, and J. Wolfe, "Vocal tract effects on the timbre of the saxophone" in SMAC13, 2013, Stockholm.

Resonances and bandwidths in the vocal tract and why they are important for speech comprehension

Noel Hanna, John Smith, and Joe Wolfe
School of Physics, University of New South Wales, Sydney NSW 2052, Australia

Abstract Summary

The bandwidths of vocal tract resonances are critical: too narrow allows speech harmonics to miss resonances, too broad gives insufficient boost to identify phonemes. We report the first measurements of bandwidths and low frequency behaviour.

Keywords: vocal tract; resonance; loss; acoustics; speech

I. INTRODUCTION

The resonances of the vocal tract act to filter the harmonically rich sound from the vocal folds to produce formants, i.e. broad peaks in the frequency spectrum. Formants characterise phonemes. Their bandwidths are important in intelligibility and perceived naturalness of the voice, and consequently are also important in speech synthesis.

Formant measurements have limited frequency resolution: frequency is sampled only at the harmonics of the voice and so gives only estimates of bandwidth. Most previous studies of resonances have used either the voice sound itself or swept sine excitation at the neck with the glottis closed. More recently broadband sound has been used to stimulate the tract from outside the neck or open mouth. Measurements made with the sound source at the neck are subject to an unknown transfer function. Broadband excitation at the open mouth cannot measure resonance bandwidth or amplitudes because of the low magnitude of the radiation impedance in parallel at the mouth.

II. METHOD

Acoustic impedance measurements were made at the lips using the three-microphone-three-calibration technique [1] and a synthesised broadband signal. Nine subjects created an air-tight seal with their lips around the measurement head, and produced the vocal gesture corresponding to the vowel sound in "heard" [ə]. Simultaneous measurements of the frequency, bandwidth and magnitude of the vocal tract resonances were made between 10 and 4200 Hz with and without phonation.

III. RESULTS

Figure 1 shows one vocal tract impedance spectrum measured without phonation and the glottis closed. All but the lowest two extrema are qualitatively similar to those of a rigid, closed cylinder of comparable dimensions, but their Q factors suggest losses several times larger than those of the rigid cylinder.

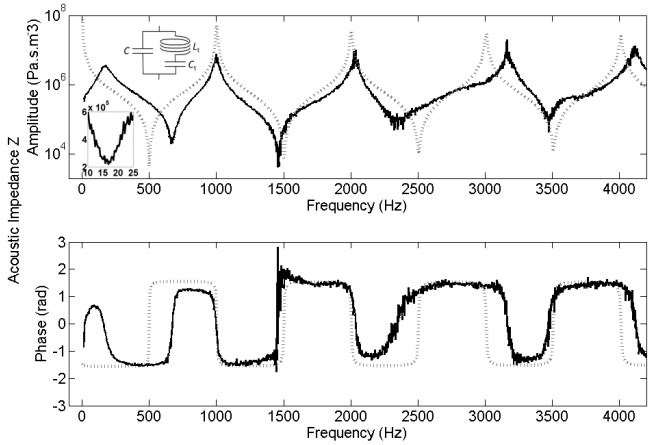


Figure 1 Impedance spectrum measured at the lips with glottis closed (solid) and, for comparison, a rigid cylinder (dashed). One inset shows the low frequency behaviour. The other shows the low-frequency, compact-object equivalent circuit.

For all subjects, the magnitudes of impedance maxima and minima are in the range 20-100 kPa.s.m^{-3} and 2-10 MPa.s.m^{-3} respectively. The bandwidths measured with a closed glottis are typically around 50 Hz for resonances and antiresonances between 400 Hz and 3 kHz and increase outside this range.

Below 300 Hz the vocal tract is not rigid. The air in the tract acts as an acoustic compliance C (*cf*. spring) of approximately 7x10^{-10} m^{3}Pa^{-1}, and its walls can be modelled as an acoustic inertance L_t (*cf*. mass). From the parallel resonance of these, which corresponds to the first impedance maximum, the inertance is estimated as 900 kg.m^{-4}, which gives a wall thickness of ~1 cm. The walls are also supported by tissue elasticity. The series resonance of this tissue compliance C_t and L_t produces the first impedance minimum, whose frequency gives C_t ~7x10^{-8} m^{3}Pa^{-1}, which gives a spring constant of ~10 N.cm^{-1} for a wall thickness of 1 cm.

REFERENCES

[1] Dickens, P., J. Smith, and J. Wolfe, *Improved precision in measurements of acoustic impedance spectra using resonance-free calibration loads and controlled error distribution*. J. Acoust. Soc. America, 2007. 121, 1471-1481.

MEASUREMENTS OF THE AERO-ACOUSTIC PROPERTIES OF THE VOCAL FOLDS AND VOCAL TRACT DURING PHONATION INTO CONTROLLED ACOUSTIC LOADS

Noel Hanna, John Smith, and Joe Wolfe

International Congress on Acoustics, Montreal, Canada, 2013

INTRODUCTION

The aero-acoustic properties of the vocal folds and tract are difficult to measure directly. Here, they were measured using broad- and narrow-band excitation at the mouth during phonation into various acoustic loads, including a non-resonant load provided by an acoustically infinite waveguide with cross section comparable with that of the tract. The tract is treated as a duct terminated by the larynx. Mechanical properties of the walls and terminations were determined using a three microphone head (Dickens *et al.* 2007).

The vocal fold response was monitored with an electroglottograph and wall motion was measured electromechanically. The impedance spectra show negative resistance bands at frequencies near those of phonation, consistent with regeneration at the folds. The measurements indicate that the walls have inertances consistent with thicknesses of order 1 cm and compliances consistent with distributed stiffnesses of about 100 kN/m^3 (Hanna *et al.* 2012). The duct resonant properties are consistent with losses several times higher than the viscothermal losses at smooth rigid walls.

METHOD

Vocal Tract Properties: Three Microphone, Three Calibration Broadband Measurements

Acoustic impedance measurements were made at the lips using a three-microphone array calibrated with three non-resonant loads (Dickens *et al.* 2007). Seven male subjects made an air-tight seal with their lips around the measurement head, and were asked to either produce the long vowel sound in the word "heard" [ɘ:], or the same vocal gesture without phonation with an open or closed glottis. Simultaneous measurements of the frequency, bandwidth and magnitude of the vocal tract resonances were made with a frequency resolution of ~2.7 Hz from 14 to 4200 Hz. A typical single cycle of the measurement lasting 370 ms is shown in Figure 1. The average of similar cycles during the same vocal gesture was used to determine the bandwidths of the resonances.

Vocal Tract Properties: Non-Rigid Behaviour

Low frequency broadband measurements were also made for each subject. However, to improve signal:noise ratio, narrower frequency bands were used (10-50 Hz at ~0.3 Hz resolution and 14-400 Hz at ~0.7 Hz resolution). In addition, a small magnet was attached to the cheek and/or neck of four of the subjects. This allowed us to measure tissue velocity by recording the Faraday EMF in a coil of wire at a fixed distance from the magnet.

For semi-quantitative analysis of the tissue motion, Figure 2b shows simple physical and electrical models of the vocal tract behaviour that are appropriate at very low frequencies, at which the tract can be approximated as a compact object (i.e. the dimensions are much less than the wavelength of sound so the pressure is approximately uniform). Electrical analogues such as these have been proposed previously (e.g. Fant 1972, Ishizaka *et al.* 1975), however we know of no measurements of the frequencies and bandwidths of the LC resonances.

Vocal Fold Properties: Three Microphone Three Calibration Narrow Band Measurements

In addition to the broadband impedance measurements during phonation, the motion of the vocal folds was probed using narrow band excitation around the frequency of phonation, while subjects phonated in the upper and lower ranges of their normal voice (Mechanism 1). A quasi-infinite pipe (length ~200 m) behind the measurement head provided a purely resistive acoustic load on the vocal folds. The vocal fold behaviour was monitored with an electroglottograph.

RESULTS

The broadband impedance spectrum in Figure 1 shows that all but the lowest two impedance extrema of the vocal tract are qualitatively similar to those of a rigid, closed cylinder of comparable dimensions. However, the Q factors of the vocal tract correspond to losses several times larger than those of a rigid cylinder with purely visco-thermal losses, in line with previously reported data on a single subject (Hanna *et al.* 2012).

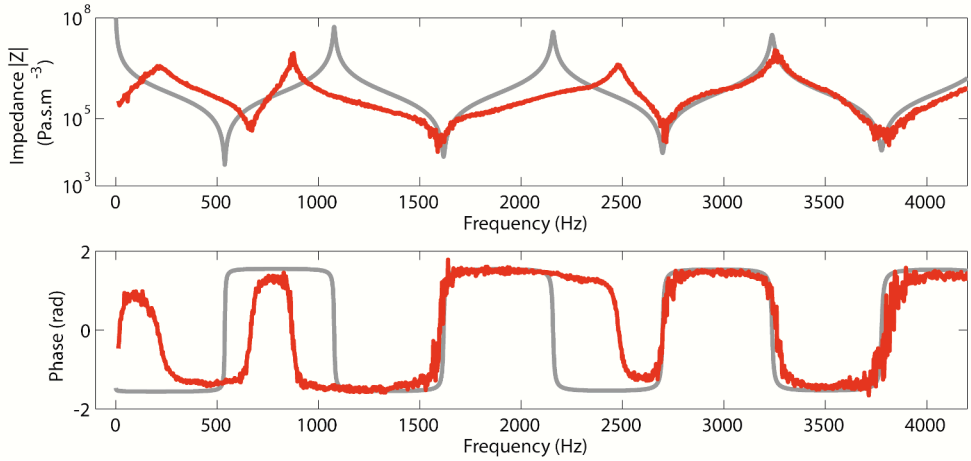


FIGURE 1. The dark curve shows a typical impedance spectrum obtained with the glottis closed for a single subject. Several cycles of such measurements are made lasting 370 ms each. The average of similar cycles during the same vocal gesture is then used for analysis. The pale line shows the calculated impedance spectrum of a rigid duct of similar dimensions (Fletcher & Rossing 1991).

At frequencies below several hundred hertz the vocal tract impedance deviates from that of a rigid tube. A low frequency broadband impedance curve is shown in Figure 2a. This non-rigid behaviour at low frequencies can be simply modelled as the compliance of the air in the tract in series with the inertance and compliance of the vocal tract walls, as shown by the equivalent circuits in Figure 2b. The measured frequencies give inertances consistent with a wall thickness of order 1 cm and compliances consistent with a distributed stiffness of about 100 kN/m^3 (Hanna *et al.* 2012).

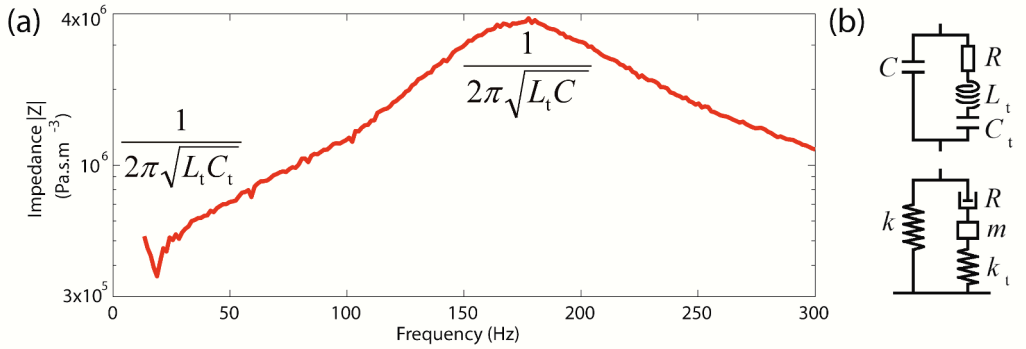


FIGURE 2. (a) Low frequency behaviour of the vocal tract. (b) Physical and electrical analogues. At ~200 Hz the volume of air in the vocal tract acts approximately as a compact acoustical element, here a spring of constant k (compliance C) in parallel with the mass, m (inertance L_t) of the vocal tract walls, this gives rise to a resonance with the observed impedance maximum. At lower frequencies k is an open circuit, and a series resonance between the mass of the vocal tract walls and their own spring k_t (compliance C_t) gives rise to an impedance minimum at ~20 Hz. Mechanical losses that lead to the finite Q of the resonances are approximated as a resistance R.

The non-rigid behaviour of the vocal tract was also observed in the motion of the magnet placed on the cheek and/or neck of four of the subjects. Peaks in the mechanical admittance were measured corresponding to a large amplitude vibration of the cheeks, as shown in Figure 3, and to a lesser extent of the neck, similar to the pattern measured by Fant *et al.* (1976).

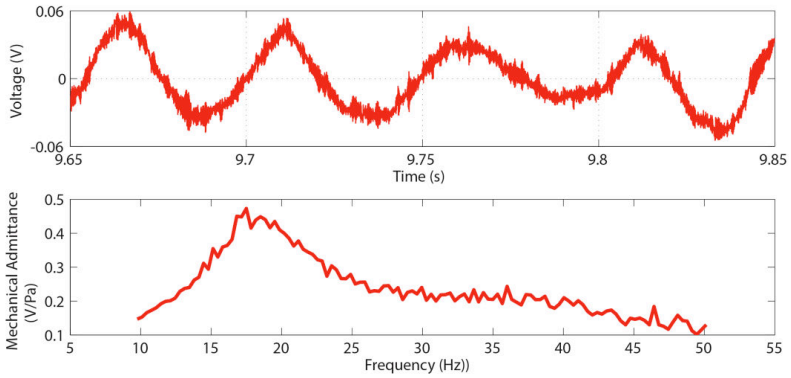


FIGURE 3. The EMF induced in a fixed coil of wire by a magnet placed on the subject's cheek during phonation. The tract was excited by broadband sound between 10 and 50 Hz. The peak in the mechanical admittance (large amplitude vibration) corresponds to the frequency (~20 Hz) of the first impedance minimum of the vocal tract.

The frequency of phonation, f_0 and its harmonics appear in the impedance spectra as additional signals with a phase uncorrelated with the synthesised sine waves of the probe signal (Figure 4). Bands of negative resistance appear close to these frequencies, consistent with regeneration at the folds.

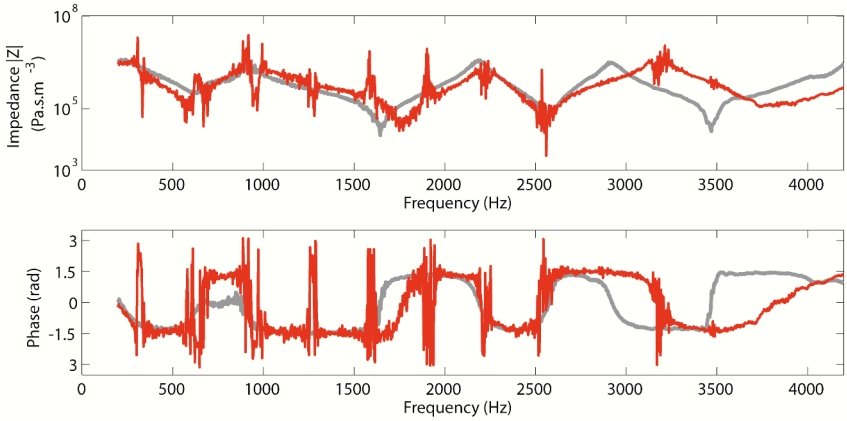


FIGURE 4. The mean impedance spectrum magnitude and phase of several cycles during phonation clearly show the fundamental frequency f_0 and harmonics of the voice. The smooth pale line shows a measurement during exhalation, i.e. with the glottis open but no phonation.

ACKNOWLEDGMENTS

We thank the Australian Research Council for its support and our volunteer subjects. The presentation of this paper was supported by a travel award from the New South Wales division of the Australian Acoustical Society.

REFERENCES

Dickens, P., Smith, J., Wolfe, J., **(2007)**. "Improved precision in measurements of acoustic impedance spectra using resonance-free calibration loads and controlled error distribution," J. Acoust. Soc. Am. **121**, 1471-1481.

Fant, G., **(1972)**. "Vocal tract wall effects, losses, and resonance bandwidths," STL-QPSR **2-3**: 28-52.

Fant, G., Nord, L., Branderud, P., **(1976)**. "A note on the vocal tract wall impedance." STL-QPSR **4**, 13-20.

Fletcher, N. H., Rossing, T. D., **(1991)**. *The physics of musical instruments* (Springer-Verlag, New York), Chap. **8**, pp. 175-182.

Hanna, N., Smith, J., Wolfe, J., **(2012)**. "Low frequency response of the vocal tract: acoustic and mechanical resonances and their losses," Proc. Australian Acoust. Soc.

Ishizaka, K., J. French, Flanagan, J., **(1975)**. "Direct determination of vocal tract wall impedance." Acoustics, Speech and Signal Processing, IEEE Transactions on **23**(4): 370-373.

SINGING EXCISED HUMAN LARYNGES: RELATIONSHIP BETWEEN SUBGLOTTAL PRESSURE AND FUNDAMENTAL FREQUENCY

N. Hanna[1,2], N. Henrich[1], A. Mancini[3], T. Legou[4], X. Laval[1], P. Chaffanjon[1,3]

[1] Department of Speech and Cognition, GIPSA-lab, Université de Grenoble, France, noel.hanna@gipsa-lab.grenoble-inp.fr
[2] School of Physics, The University of New South Wales, Australia
[3] Laboratoire d'Anatomie Des Alpes Françaises, Faculté de Médecine de Grenoble, France
[4] Laboratoire Parole et Langage, Aix-en-Provence, France

***Abstract: Ex vivo* studies of excised human larynges allow measurements of physical parameters that are important for our understanding of speech and cannot be otherwise determined from *in vivo* studies. This study focuses on the relationship between subglottal pressure and fundamental frequency of phonation, under several experimental conditions: two different air supply (pressure or flow), and three different control settings (pressure or flow, arytenoid compression, vocal-fold extension). A female human larynx was used.**

The results demonstrate a general linear behavior between subglottal pressure and fundamental frequency, with two different gradient regions (50 Hz/hPa or 1 Hz/hPa) depending on the range of downstream pressure. The nonlinear behavior of the human laryngeal system is well known and is illustrated here by pitch jumps and hysteresis cycles. The driving flow source enables the activation of a high-frequency vibratory mode between 600 and 1000 Hz, and greater dynamic ranges of subglottal pressure and sound pressure level.

Keywords : **Excised human larynges, pressure source, flow source, fundamental frequency, subglottal pressure, nonlinear laryngeal behavior**

I. INTRODUCTION

Human voice production is controlled by airflow supply interacting with laryngeal and vocal-tract configurations. Airflow provides an aerodynamic energy to the vocal system, through subglottal pressure and flow. The conversion of aerodynamic energy to mechanical and acoustical energy depends strongly on the biomechanical properties of the vocal folds, which is characterized by the fundamental frequency (f_0) of vibration. Subglottal pressure and f_0 are therefore two key source parameters in voice production.

The relationship between subglottal pressure and f_0 has been studied theoretically, e.g. [1], and experimentally (for a review, see [2]). Fundamental frequency typically rises from 1 to 7 Hz for an increase in pressure of 1 hPa [1,2], but values as high as 20 Hz/hPa can be found [2,3]. The relationship is nonlinear and it depends on the degree of vocal-fold elongation and adduction [2].

Over the past fifteen years, the subglottal pressure-f_0 relationship has been experimentally studied mainly on excised larynges [4], for the sake of a better experimental control of vocal-fold settings and due to the fact that direct measurement of subglottal pressure in humans is very invasive.

For reasons of practicality, excised larynx experiments are often performed on species other than humans [5]. Even if these non-human larynges may be used for assessing aerodynamic and acoustical models, they greatly differ with respect to biomechanical properties of the vocal folds, to laryngeal anatomy and to pressure-frequency behavior [5]. Furthermore, studies of excised dog, pig, sheep and cow larynges have noted that the supra-glottal tissues and structures in some animal models may cause a non-zero supra-glottal pressure, which may have an effect on the glottal resistance [6,7].

Recent measurements of phonation threshold pressure and flow on excised human larynges within 24 hours postmortem reported that phonation threshold pressures were typically lower than those measured in excised canine [8].

The relationship between subglottal pressure and f_0 in human voice production remains an open question. This preliminary study aims first to explore this relationship in excised human larynges. A second aim is to explore the impact of the type of airflow supply on the pressure-frequency relationship. Indeed, a constant flow source is a more physiological boundary condition below the glottis, which may impact voicing [9].

II. METHODS

This paper presents measurements made on a human female larynx, which was harvested and frozen within 48 hours postmortem, and thawed before the experiment. The larynx was made to self-oscillate by application of a source of pressurized air.

The specimen was secured to the experimental bench posteriorly and anteriorly at the cricoid cartilage. An

intubation tube was used to supply the source of pressurized air.

Crico-thyroid tension was maintained throughout the experiment by securing the epiglottis. The aryepiglottic folds were held separate throughout by the use of Beckman-Eaton laminectomy retractor.

In order to simulate the lateral crico-arytenoid muscle action used *in vivo*, the level of adduction/abduction of the vocal folds was modified by manual control of metal probes inserted into the muscular processes of the arytenoid cartilages.

The oscillation of the vocal folds was monitored by an Electroglottograph (EGG) (EG2-PCX2, Glottal Enterprises, Syracuse, New York, USA) with electrodes aligned with the plane of the vocal folds. The fundamental frequency of vibration (f_0) was estimated from the EGG signal using the YIN algorithm [10].

Subglottal pressure was measured by means of a specialized aerodynamic workstation (EVA2, S.Q.Lab, Aix-en-Provence, France) with a probe inserted into the crico-thyroid ligament.

A microphone was placed at a distance of 30 cm from the larynx in order to monitor the experiment and measure the sound pressure level (SPL).

The experimental setup is illustrated in Fig. 1.

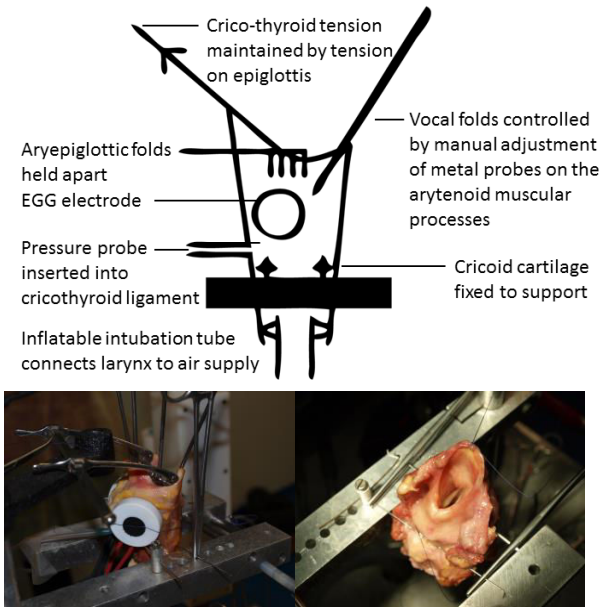


Figure 1. Schematic description of the experimental setup (top) and photographic views (bottom).

The study was conducted in two parts. In the first part (sequence 1) the oscillation in the larynx was driven by a pressure source (Mecafer, LT50 Compressor). The driving pressure was first varied while the larynx geometry was kept fixed (1A). It was then kept constant while either the lateral compression of the arytenoid cartilages was varied (1B), or the vocal fold extension was varied (1C). In the second part of the study (sequence 2), the oscillation was driven by a flow source. The driving flow was first varied (2A). It was then kept constant while either arytenoid compression (2B) or vocal fold extension (2C) were varied. The independent variables are summarized in Table 1.

Table 1. Experimental sequences

Sequence Number	Source Type	Independent Variable
1A	Pressure	Subglottal pressure
1B	Pressure	Compression of arytenoids
1C	Pressure	Vocal fold extension
2A	Flow	Subglottal flow
2B	Flow	Compression of arytenoids
2C	Flow	Vocal fold extension

III. Results

Larynx behavior

As shown in Fig. 2, the human female larynx was able to oscillate over a wide range of fundamental frequency, typically between 300 and 900 Hz. The corresponding dynamic range of SPL was approximately 40 dB, reaching a maximum of 90 dB at 30 cm. Subglottal pressures varied up to 40 hPa.

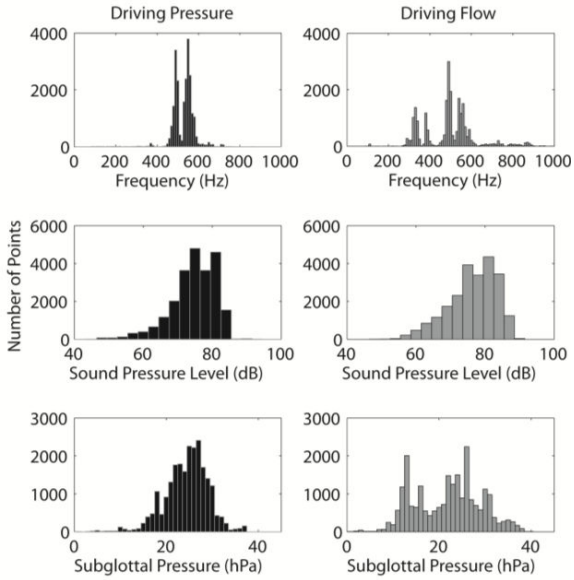


Figure 2. Histograms of fundamental frequency (top panels), SPL (middle panels), and subglottal pressure (bottom panels). Either a driving pressure source (left panels) or a flow source (right panels) were used. The bin sizes are 10 Hz for frequency, 3 dB for SPL and 1 hPa for subglottal pressure.

Fig. 3 plots the relationship between subglottal pressure and f_0 across all experimental conditions listed in Table 1. The results for the driving pressure source are shown in black, and the driving flow source are in grey.

When the larynx configuration is kept fixed (Fig. 3(A)), two distinct regions are highlighted with solid black lines showing the gradients. For subglottal pressure lower than 10-15 hPa, an increase of about 50 Hz/hPa is observed, for oscillatory frequencies f_0 between ~100 and 500 Hz (~A2-B4). For subglottal pressure higher than 10 hPa, the relationship shows a gradient of ~1 Hz/hPa.

A similar gradient of ~1 Hz/hPa is found when the laryngeal configuration is varied (Fig. 3(B) and 3(C)). Differences between pressure source and flow source are observed, which will be discussed later in the paper.

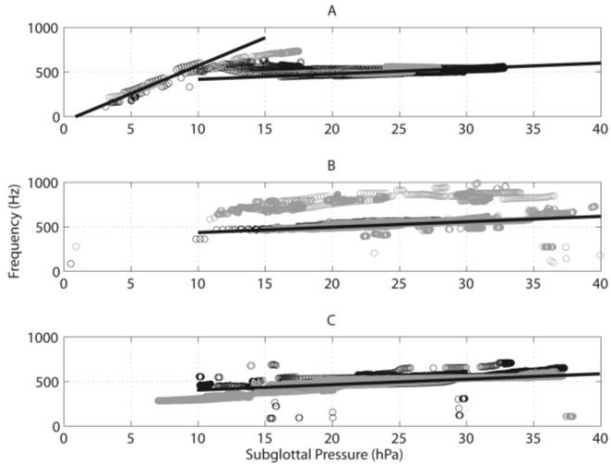


Figure 3. Relationship between subglottal pressure and f_0, with (A) variation of flow (grey) and pressure (black), (B) variation of arytenoid compression under constant flow (grey) and pressure (black), (C) variation of vocal fold extension under constant flow (grey) and pressure (black). Straight lines show approximate gradients.

Vocal fold nonlinearity

The variation of driving source (A), arytenoid compression (B) and vocal fold extension (C) all show a globally linear relationship between subglottal pressure and f_0 as shown by the straight lines in Fig. 3. However, the relationship is locally nonlinear, as it is interrupted by sudden pitch jumps. The pitch jumps were observed under all experimental conditions, as illustrated by samples in Fig. 4.

Pitch jumps occurred between 20 and 30 hPa, with a variability which may reflect small changes in the control parameters. A hysteresis effect is clearly demonstrated in each case, due to the nonlinear behavior of the laryngeal vibratory system. The gradients between subglottal pressure and f_0 calculated on either side of the jumps are of similar values.

In general, upward jumps occurred at higher frequencies than downward jumps. However, in some cases for the driving pressure source condition, the downward pitch jumps occurred at a higher frequency that the upward jump, as shown in Fig. 4(A).

Fig. 4(C) shows two successive pitch jumps under the driving flow source condition. A second pitch jump is observed above 30 hPa, a result which was also found under the constant pressure condition for similar elongation of the vocal folds.

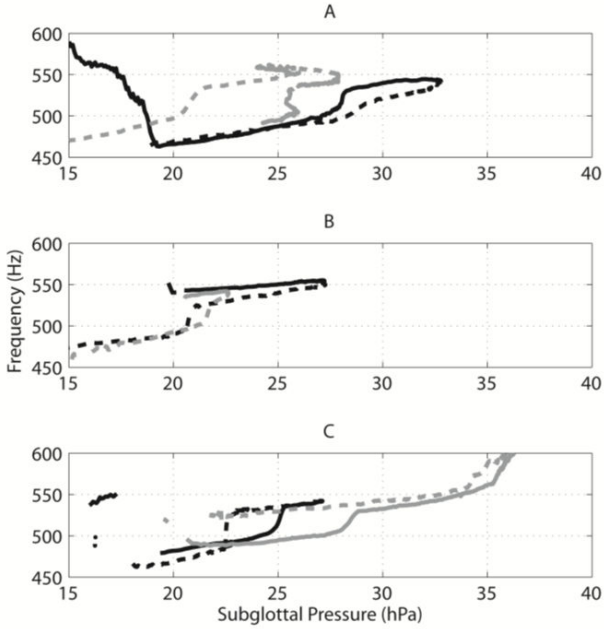


Figure 4. Examples of nonlinear behaviors and pitch jumps with (A) variation of flow (grey) or pressure (black), (B) variation of arytenoid compression under constant flow (grey) and pressure (black), (C) variation of vocal fold extension under constant flow (grey) and pressure (black). Solid lines show increases in the independent variable, dashed lines show decreases.

Driving pressure and driving flow sources

Fig. 2 illustrates two main differences between the experimental conditions. The first is that the flow source allows oscillation at lower subglottal pressures, with greater sound pressure level at frequencies in the normal range of phonation (<500 Hz). Furthermore as shown in Fig. 3(C) there is a slightly steeper gradient in the relationship between subglottal pressure and f_0 at these frequencies.

Secondly, the flow source activates an extreme high frequency mode of oscillation between 600 and 1000 Hz, as can be seen in Fig. 3(B) while the arytenoid compression was varied. In contrast to the behavior under the constant pressure condition (shown in black), two different modes of vibration are found at the same

subglottal pressure. The fundamental frequency of oscillation can occur either between ~500 and 600 Hz comparable with the pressure source case, or ~600-1000 Hz (musically ~D5 – C6), frequencies typically achieved only by sopranos.

This latter 'soprano' mode of oscillation was highly unstable and occurred in 15% of measurements with the flow source. Fundamental frequencies above 600 Hz were also observed in 5% of measurements with the pressure source but oscillation did not occur higher than 800 Hz.

These high fundamental frequencies were only observed towards the end of the experimental session. They are therefore possibly due to dehydration of the tissue over time and as such may not reflect the normal range of oscillation of the vocal folds.

IV. Discussion

In all experimental conditions a similar linear increase of f_0 with subglottal pressure above 10 hPa or 500 Hz with a gradient of ~1 Hz/hPa was observed in line with previous measurements [1,2]. In this region, there were typically one or two pitch jumps of ~40 Hz, which showed hysteresis with increasing and decreasing of the independent control parameter, similar to observations of nonlinearity under variation of vocal fold tension [11].

When the larynx geometry was fixed and the pressure or flow was varied, a steeper gradient of ~50 Hz/hPa was observed below 15 hPa, more than twice the values found in the literature [2,3]. However, the experimental data is scarce in this region, and it appears that there may be a difference in behavior over this range under the driving flow condition. The understanding of the behavior in this region calls for further research.

V. Conclusion

These preliminary data produced from a study of one female human larynx highlight differences in vocal fold behavior due to changes in the subglottal air supply.

Further work aims to improve the experimental setup to allow for quantifiable measurements of arytenoid movements and to extend the process to allow measurement of the downstream air flow.

References

[1] I.R. Titze, "On the relation between subglottal pressure and fundamental frequency in phonation," *J. Acoust. Soc. Am.*, vol. 85(2), pp. 901-906, 1989.

[2] F. Alipour, and R.C. Scherer, "On pressure-frequency relations in the excised larynx," *J. Acoust. Soc. Am.*, vol. 122(4), pp. 2296-2305, 2007.

[3] P. Lieberman, R. Knudson, and J. Mead, "Determination of the rate of change of fundamental frequency with respect to subglottal air pressure during sustained phonation," *J. Acoust. Soc. Am.*, vol. 45(6), pp. 1537–1543, 1969.

[4] J. Van den Berg, and T. Tan, "Results of experiments with human larynxes," *ORL*, vol. 21(6). pp. 425-450, 1959

[5] F. Alipour, and S. Jaiswal, "Phonatory characteristics of excised pig, sheep, and cow larynges,". *J. Acoust. Soc. Am.*, vol. 123 (6), pp. 4572–4581, 2008.

[6] F. Alipour, and S. Jaiswal, and E. Finnegan, "Aerodynamic and acoustic effects of false vocal folds and epiglottis in excised larynx models," *The Annals of otology, rhinology, and laryngology*, vol. 116(2), pp. 135-144, 2007.

[7] Alipour, F. and S. Jaiswal (2009). "Glottal airflow resistance in excised pig, sheep, and cow larynges." *Journal of Voice*, vol. 23(1), pp. 40-50.

[8] T. Mau, J. Muhlestein, S. Callahan, K.T. Weinheimer, and R.W. Chan, "Phonation threshold pressure and flow in excised human larynges," *The Laryngoscope*, vol. 121(8), pp. 1743-1751, 2011.

[9] M.S. Howe, and R.S. McGowan, "Voicing produced by a constant velocity lung source,", *J. Acoust. Soc. Am.*, vol. 133(4), pp. 2340–2349, 2013.

[10] A. De Cheveigné, and H. Kawahara, "YIN, a fundamental frequency estimator for speech and music," *J. Acoust. Soc. Am.*, vol. 111(4), pp. 1917-1930, 2002.

[11] J. G. Švec, H. K. Schutte, and D.G. Miller, "On pitch jumps between chest and falsetto registers in voice: Data from living and excised human larynges," *J. Acoust. Soc. Am.*, vol. 106(3), pp. 1523-1531, 1999.