Technical University of Denmark

Jesper Pedersen (s112357)

The singing voice: A vibroacoustic problem

Master's Thesis, August 2013

DTU Electrical Engineering
Department of Electrical Engineering

JESPER PEDERSEN (S112357)

The singing voice: A vibroacoustic problem

Master's Thesis, August 2013

Supervisors:

Finn T. Agerkvist

DTU - Technical University of Denmark, Kgs. Lyngby - 2013

The singing voice: A vibroacoustic problem

This report was prepared by:
Jesper Pedersen (s112357)

Advisors:
Finn T. Agerkvist

Project period: February 2013- August 2013

ECTS: 35

Education: MSc

Field: Acosutical Engineering

Table of Contents

Abstract

Although the human voice have been under the microscope for many decades, its workings are still debated by scientists, doctors and voice specialists. Complete Vocal Technique is a singing techniques that strives to cover all the sounds and effects that the voice can do in a musical situation. The main focus of the project was to uncover the underlying mechanisms of this particular singing technique and its characterization of the singing voice in four modes.

Through a study of the anatomy and physiology of the singing voice, the basic knowledge for approaching this problem was acquired. The technique could then be analysed for clues about the key factors, and it was stated that one of them was the aero-acoustic coupling between the vocal tract and the vocal folds. To analyse this hypothesis, data describing the geometry of the vocal tract was acquired through MRI scanning, which resulted in a set of simple one dimensional area function of the vocal tract, describing the cross-section from the vocal folds to the lips. The extraction of this area function was partly validated through simulation and experiments. After the process of acquiring the data for the model, little time was left for the actual study of the coupling. A preliminary study of the vocal tract impedance was performed and a simple model of the vibration and flow through the vocal folds coupled to the vocal tract was performed. The model did however not yield the results that was anticipated by acoustical measurements, and unfortunately the project did not leave time for more iterations of updating the model.

Preface

This report is the documentation of a master thesis in Engineering Acoustics at the Danish Technical University (DTU). The project stretched from February 2013 to August 2013 under the supervision of Assoc. Prof. Finn Agerkvist. The study was on modelling the vibro-acoustics of the singing voice.

The motivation for the project was to investigate some of the underlying mechanisms of the art of singing. Specifically the singing technique, Complete Vocal Technique, which defines a technique where the singing voice is characterize into four modes. The characterization of these modes were the main focus of this project. For the benefit of the singer understanding his/her instrument.

The project includes an MRI study, which was made in collaboration with Hvidovre Hospital and Danish Research Centre for Magnetic Resonance (DRCMR).

A special thanks to Lars Hanson, Cathrine Sadolin, Henrik Kjelin and especially thanks to test subjects Kaare Thøgersen and Finn Agerkvist.

Jesper Pedersen (s112357)

CHAPTER 1

Introduction

The human voice makes for a very versatile musical instrument with its wide range of timbres and a pitch range of up to about four octaves. On top of that, it is relatively easy to learn to use and almost everybody has one.

In the search for an understanding of the physiology of this system, much of the research has been conducted through studies of the human voice in speaking conditions. This approach covers the general functionality, but the versatility of the system leaves many conditions that calls for further investigation of the physiology.

Prior studies have been conducted on the singing voice, but the wide variety of singing styles, and the techniques that go along with them, require an even deeper understanding of the complex physiology involved in singing. The sounds of various singing styles differ tremendously even though the sounds are produced using the same system, which suggests that the mechanism of phonation may be used in different ways throughout the styles.

One of the purposes of this research is to aid the singer in understanding what he or she is trying to do, when learning these singing techniques. It is hard to imagine what goes on inside the throat while singing, but a knowledge of the physics may give some singers a reference point, from which to adjust their perception of singing.

Among the many schools of singing technique, Complete Vocal Technique (CVT) is developed to be used for all styles of singing. The technique is described in the textbook Complete Vocal Technique [Sadolin, 2012], where they have defined concepts and terminologies that will be analysed in this project.

To summarise CVT - it builds on fundamentals such as having well-established breath support and avoiding constrictions of the throat while singing. At the core of CVT is the definition of four different modes of singing, which will be the main focus of this project. These modes are:

- Neutral (non-metallic) - soft singing
- Curbing (half-metallic) - resembles a cry
- Overdrive (full-metallic) - resembles a shout
- Edge (full-metallic) - resembles a scream

The modes differ in timbre, loudness, vocal range and which vowels they can be produced on. The concept of how "metallic" a mode is, is a definition invented in CVT that describes the perception of the sound. It is believed to be closely related to how rich in overtones the sound is [Selamtzis, 2011].

The goal of this project is to model the vibro-acoustics of the singing voice, when using techniques taught in Complete Vocal Technique.

This first chapter of this report contains a familiarisation with the anatomy of the system, a literature study explaining the concepts in the physiology and finally a preliminary analysis of CVT.

In the second chapter the modelling procedure is set up for a model that is capable of simulating the phonatory system using the CVT modes.

The third chapter describes acquisition of input data for the model, via MRI scans of the vocal tract. The scans were performed as part of this project. An analysis of the acquired data on the vocal tract is performed.

The fourth contains an experimental evaluation of the data that was acquired through the MRI scans

In the fifth chapter the modelling of the system is performed - resulting in a discussion of the results

1.1 Anatomy

In the following section, an overview of the anatomy of the human phonatory system is presented. The purpose of this section is to familiarize the reader with the system. It is solely based on the following references. [GetBodySmart.com, 2013] [Rosen and Simpson, 2008] [Sundberg, 1987].

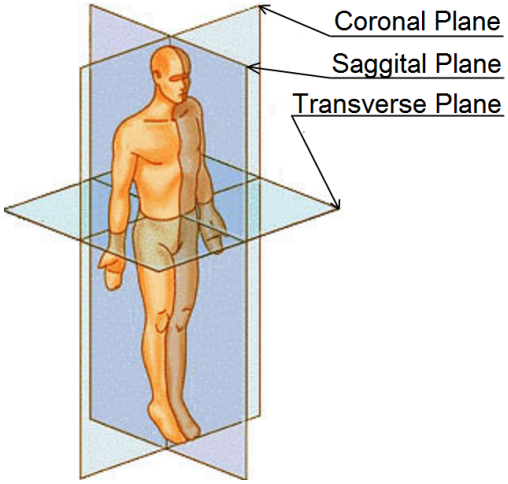


Figure 1.1: Overview of body plane terminology, adapted from [Wikipedia, 2012].

Many of the illustrations in this section show the anatomy of the phonatory system in slice planes of the body. The orientation of these planes underlies a certain medical terminology, which is explained in figure 1.1. Also: posterior means from the front, anterior - from the back, superior - from above, and inferior - from below.

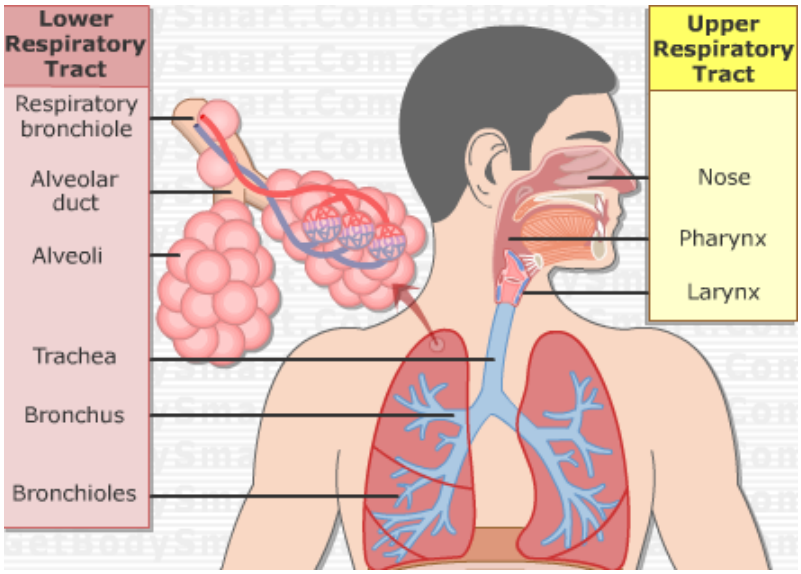


Figure 1.2: Overview of the anatomy of the respiratory system, adapted from [GetBodySmart.com, 2013].

As illustrated in figure 1.2, the human phonatory system consists of the lungs, *trachea*, *larynx*, *pharynx* and the *oral* and *nasal* cavity. Respiration is the primary function of this system, but sounds are produced when air passes constrictions as it is being pushed out by the lungs. In voiced phonation the constriction is mainly at the vocal folds, but consonants are mainly produced by constrictions in the *oral* cavity.

Lungs and trachea

The lungs consist mostly of *alveoli*, which are small cup shaped sacs where the oxygen is extracted from the air. All the small alveoli connect into branchings of small tubes in each lung called the *bronchi* (*bronchus, bronchioles*). The *bronchi* branch into one input tube from each lung, called the primary *bronchus*. The primary *bronchus* from each lung connect into one tube, called the *trachea*, which runs to the *larynx*. The *trachea* wall is build up of c-shaped cartilaginous rings that keep the circular tube open.

Larynx

The *larynx* is a complex of cartilages that contains the vocal folds, see figure 1.3. The *larynx* sits on top of the *trachea* and leads into the *pharyngeal* cavity to the mouth and nose.

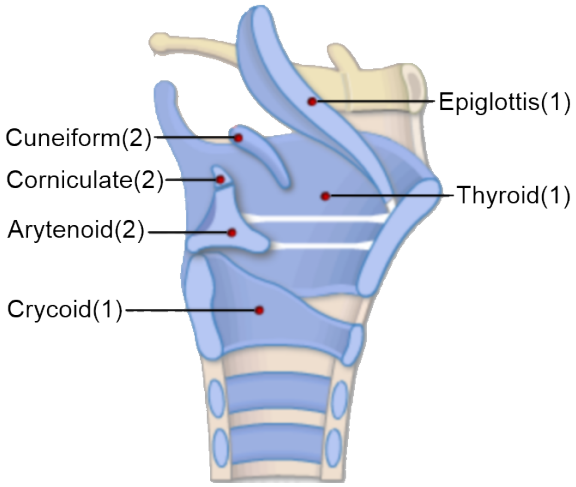


Figure 1.3: Midsaggital view of the cartilages of the larynx. The number in the parenthesis represents the number of cartilages, adapted from [GetBodySmart.com, 2013].

The *larynx* contains nine cartilages, 3 single and 3 pairs:

- The biggest is the *thyroid* cartilage that acts as a protective cover of the *larynx*, while also containing the base attachment points of the *vocalis* muscles, which are part of the vocal fold.

- The *cricoid* cartilage forms the back wall of *larynx*.
- The *epiglottis* cartilage sits in the top part of the *thyroid* cartilage and its primary function is to cover the *larynx* in the swallowing process.
- The *arytenoids* are the posterior attachment of the *vocalis* muscles. They are the structures that are positioned when adducting/abducting the vocal folds (closing/opening them).
- The *corniculate* cartilages are small pyramid-shaped nodules that provide additional points of attachment to muscles of the *arytenoids*.
- The *cuneiform* cartilages aid in adding support to the soft tissue.

The muscles that control the movement of the *larynx* cartilages are shown in figure 1.4.

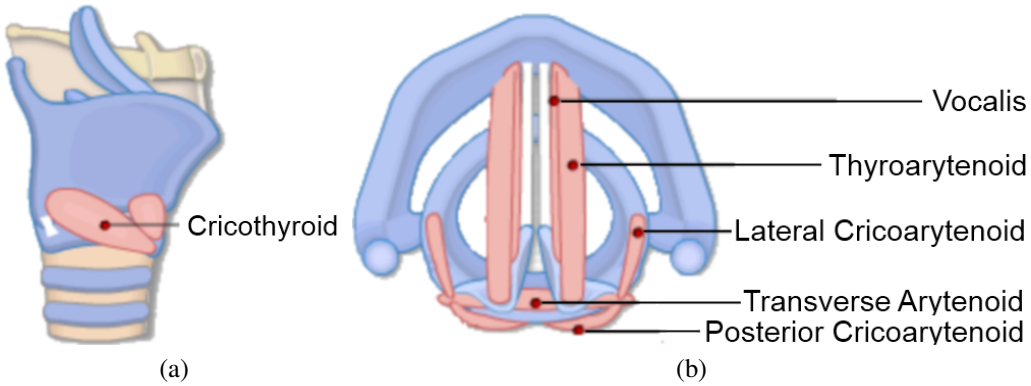


Figure 1.4: Saggital and transversal view of the intrinsic muscles of the larynx, adapted from [GetBodySmart.com, 2013].

The most important muscles involved in phonation is the pair of vocalis muscles. They are attached to the thyroid cartilage in the front and and the vocal process of the arytenoid cartilages in the back. Additional control of the voice is provided by the cricothyroid muscles and the paired posterior, lateral and transverse arytenoid muscles. Most of the muscles in the *larynx* are attached to the *arytenoids*, where they control the adduction/abduction and tension of the vocal folds by movement and rotation if tge arytenoid cartilages. The *cricothyroid* muscles provide additional stretching/slackening the vocal folds by kipping the *thyroid* cartilage.

The whole *larynx* can be raised or lowered by extrinsic muscles, which are not shown in any of the figures here. The whole structure is covered with soft tissue and a mucosa membrane that keeps the airway clean and heat up the inhaled air.

Inside the larynx is two sets of folds as seen in figure 1.5. The ventricular folds, also called the false vocal cords, are positioned above the vocal folds and consists of a thick mucosa membrane. Their primary task is to lubricate the vocal folds and help keep food out of the

airway, but they also play a role in singing and a major part in some vocal effects, such as distortion [Sadolin, 2012].

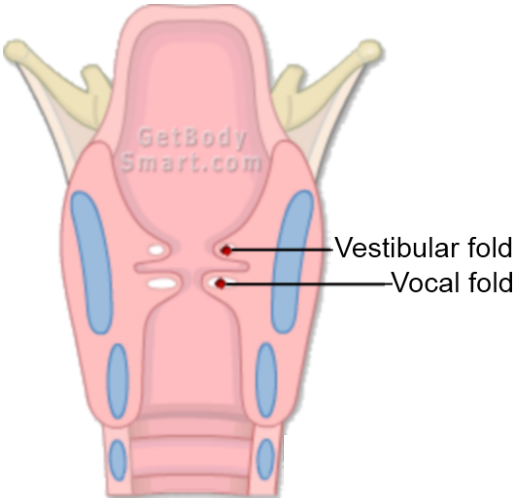


Figure 1.5: Posterior coronal view, showing the inside of the larynx, adapted from [GetBodySmart.com, 2013].

This view in figure 1.6 is seen when performing an endoscopy, which is a procedure where a camera is inserted through the nose to view the vocal folds. The gap between the vocal folds is referred to as the glottis. The *Aryepiglottal* folds and the tissue covered *epiglottis* forms a funnel above the vocal folds, called the epilarynx tube. The shape of this funnel can be controlled in singing.

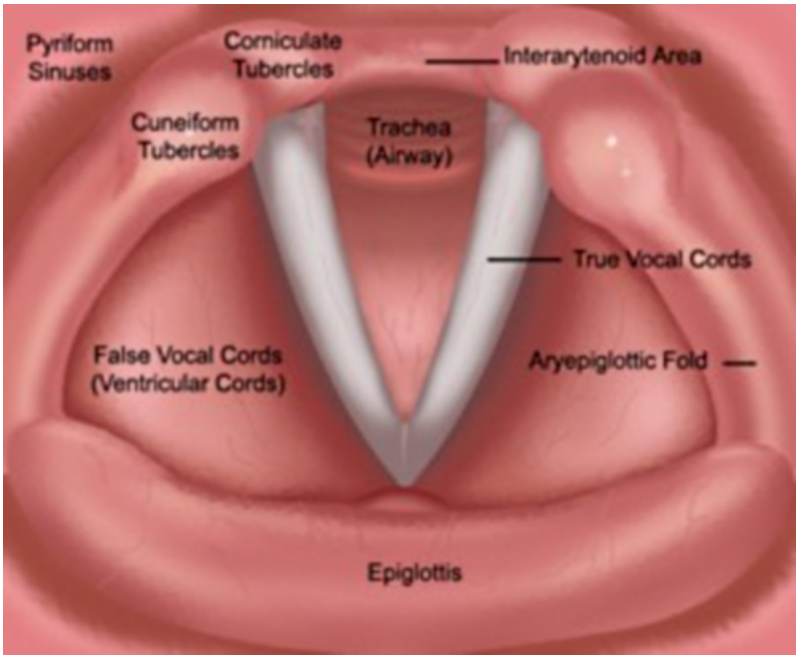


Figure 1.6: Superior view of the larynx (downwards is towards the mouth) [gbmc.org, 2013].

The vocal folds themselves are multi-layered structures containing the vocalis muscle and a cover consisting of multiple layers of different material properties, see figure 1.7.

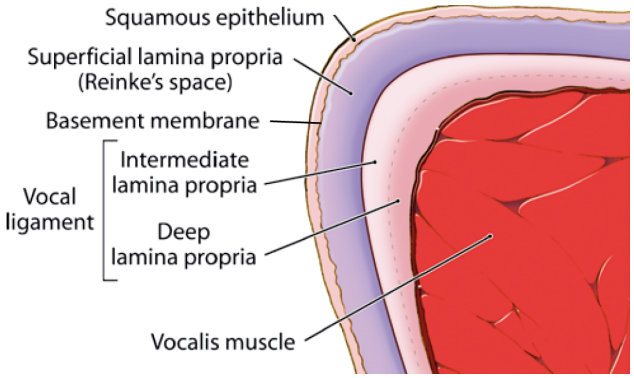


Figure 1.7: Coronal slice showing the vocal fold layers [Rosen and Simpson, 2008].

Vocal tract

The vocal tract includes the *pharynx*, *nasal* cavity and the *oral* cavity, see figure 1.8. During phonation it functions as a tube that transfers and shapes the sound from the *larynx*. The sound is shaped by varying the size of the cavities, thereby by changing the resonances of the tube. The larynx position, the epilarynx tube, pharyngeal wall and the root of the tongue can change the shape of the lower part of the vocal tract. The nasal passage can be open or closed by the soft palate to include this tube in the system. The cheeks, the lips and the tip, front and back part of the tongue shapes the cavity of the mouth. Lowest in the vocal tract on each side of the larynx are two small cavities called the piriform sinuses. These are closed up when the larynx is raised.

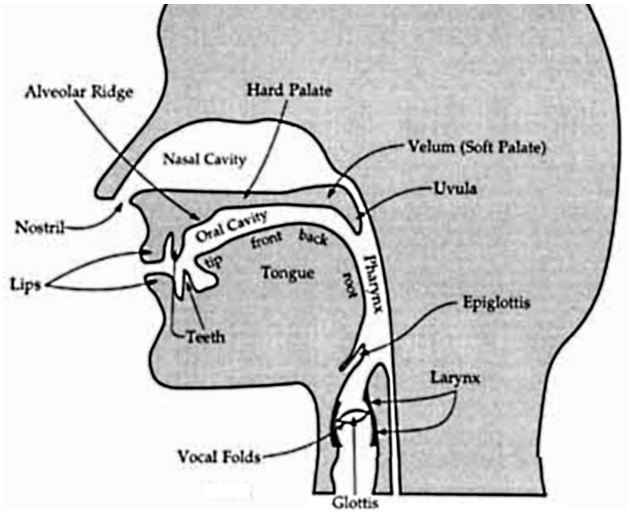


Figure 1.8: Midsaggital view of the vocal tract [Story, 1995].

1.2 Physiology

In this section the functionality of the human phonatory system is explained. The knowledge presented in this section is the result of a literature study of the physiology of singing.

The physiology of phonation is often split into two parts, the source part (vocal folds) and the filter part (vocal tract). In voiced phonation the vocal folds oscillate, creating fluctuations in the airflow from the lungs - the vocal tract then transfer these fluctuations and shapes the sound via resonances in the tract.

The source - vocal fold oscillation

The rolling movement illustrated in figure 1.9 has been observed under normal phonation. The wave that forms across the vocal fold is called the mucosal wave [Titze, 1989]. The figure shows a slice trough both vocal cords displaying their movement through one cycle of phonation.

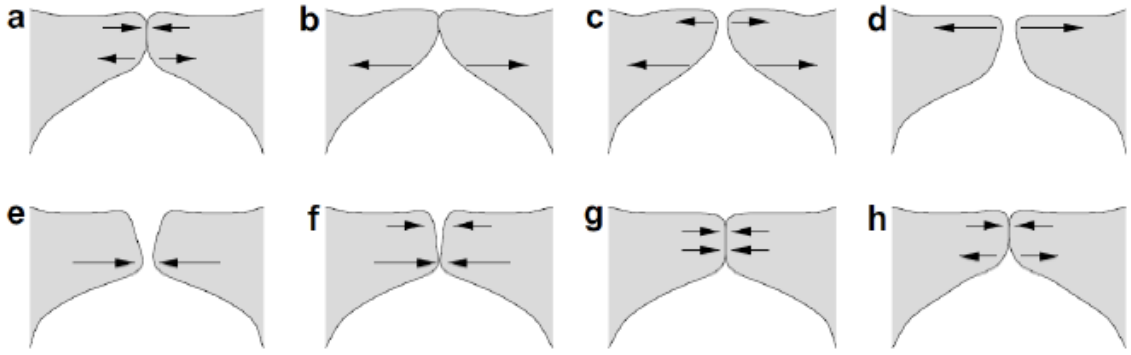
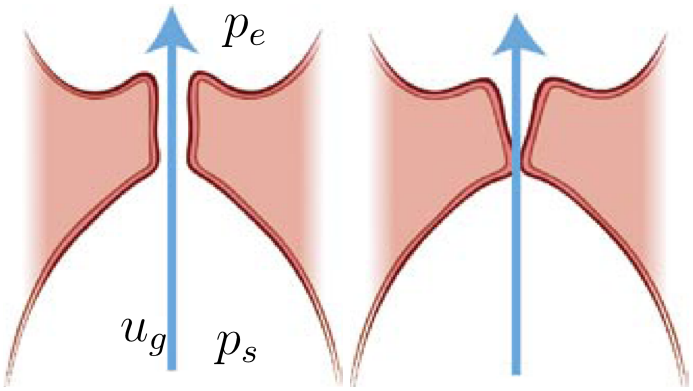


Figure 1.9: Coronal view of the vocal folds during normal phonation. Illustrating one cycle of the mucosal wave. (a-h) is chronological [Selamtzis, 2011].

The deeper layer of the vocal fold is vibrating in its first spring mode, while the cover of the vocal fold exerts the shearing motion that creates the travelling mucosal wave. Vibration roughly occurs at the resonance frequency of the string mode, which is determined mainly by the effective mass and stiffness of the vocal cord. The phonation frequency is however also affected by the collision and the aerodynamics [Sundberg, 1987].

The muscles in the larynx have to adduct the vocal folds before oscillation can start. The degree of adduction can determine whether the phonation will be breathy, normal or pressed. Breathy phonation requires loose adduction while pressed phonation requires a tight adduction [Birkholz et al., 2011].

Phonation happens when the pressure below the adducted vocal folds exceeds the threshold where the vocal folds separate. Once the vocal folds are open the air will start flowing through the glottis, see figure 1.10a. When the vocal folds have opened, the elastic forces of the vocal folds will seek to close them again, under some conditions, aided by the aerodynamic forces.



(a) Open glottis (b) Closing glottis

Figure 1.10: Aerodynamic situations in the open glottis shown on the coronal view of the vocal folds during normal phonation, Adapted from [Rosen and Simpson, 2008].

In figure 1.10, p_s is the subglottal pressure (pressure in the trachea, just below the glottis), p_e is the pressure in the epilarynx tube and u_g is the flow through the glottis. Given the mean flow and the dimensions of the vocal folds, the aerodynamics in the glottis is likely to be subject to the Bernoulli effect [Heil and Hazel, 2011]. This means that as the cross-section narrows up through the glottis, the fluid velocity will increase and the pressure will drop according to Bernoulli's equation.

$$p_e = p_s - \frac{1}{2}\rho\frac{u_g^2}{a_g^2} \tag{1.1}$$

The mean flow and dimensions of the glottis also suggests that a jet will form downstream of the narrowest cross-sectional area, detaching the flow from the walls of the glottis. This implies that Bernoullis equation is not valid in this region. The rolling motion shifts the shape of the duct through the glottis between being converging, in the opening phase and diverging in the closing phase. The resulting effect is that the jet region moves further upstream in the closing phase [Pelorson et al., 1995]. The average pressure on the glottis wall will then be larger in the opening phase than in the closing phase, enabling the passing air to excibit work on the vocal fold tissue [Heil and Hazel, 2011].

The acoustical conditions above and below the glottis have an effect of how the air flows through the glottis and how the vocal folds vibrate. The pressure difference between the

inlet and the outlet cause a flow and the pressures on either side will alter the forces that make the vocal folds oscillate. In [Titze and Worley, 2009] it is shown that the coupling between the vocal folds and the vocal tract is enhanced, if the cross-sectional area of the vocal tract just above the outlet of the glottis is small.

In [Titze, 2008] it is explained that an inertive load on the vocal folds increase the amount the work done in the vocal fold vibration. The load of the vocal tract is inertive in the frequency range below the first formant [Titze, 1988]. The compliant load just above the first formant would have the opposite effect, squelching the oscillation if the coupling is strong enough[Titze, 1988].

The pulsating flow through the glottis creates pressure waves in the vocal tract propagating towards the mouth, but pressure waves are also created in the trachea propagating towards the lungs. These create acoustic conditions that generally are enhancing phonation [Austin and Titze, 1997]. In a recent study [Sundberg et al., 2013] it is shown that fluctuations in subglottal pressure have lesser effect in pressed phonation.

The fluctuating pulse creates a harmonic signal with a fundamental frequency corresponding to the pulsation frequency, while the strength of the harmonics depend on the shape of the pulse. It has been observed that an abrupt closing phase of vocal folds lead to a harmonic signal that rolls off slower with frequency, i.e. is richer in overtones [Henrich et al., 2004]. The volume velocity of the glottal source is often approximated as being proportional to $1/f^2$ [Fulop, 2011], but this can differ depending on the condition of the source. A general spectrum of the source is found in figure 1.11.

Shaping of the sound

The flow through the glottis creates pressure waves that propagate through the vocal tract and radiates out through the lips. The waves are by some approximation plane waves in a tube, up to about 4-5 kHz, which covers most of the energy of the signal during singing [Story, 1995].

The varying cross-section of the vocal tract will generate reflected waves that travel back towards the glottis. The interaction between the forward and backward propagating waves will create standing waves, that will resonate at some frequencies.

When the waves reach the lips, part of them will be radiated out and part of them will be reflected back again. The output of the vocal tract will be a signal that has the harmonics of the source signal, but some harmonics will be amplified by the resonances, see figure 1.11. A change in the shape of the vocal tract shifts the resonances and thereby change the perceived vowel. In phonation theory the resonances are often called formants and are denote F1, F2, etc., when analysing the voice signal [Sundberg, 1987].

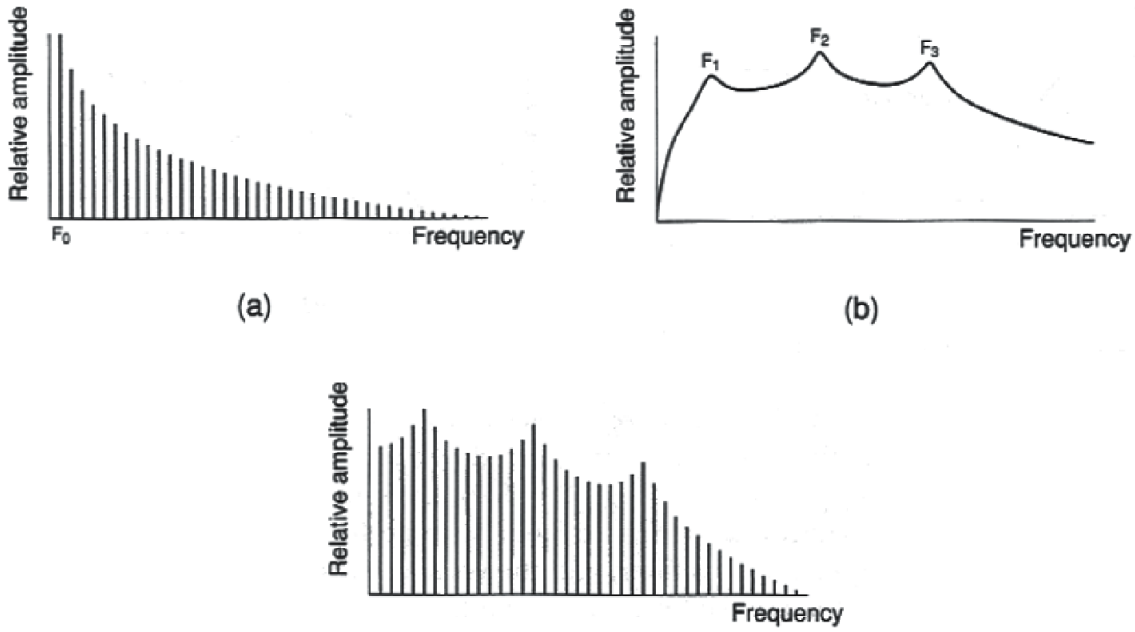


Figure 1.11: The source-filter effect for the vocal tract [Selamtzis, 2011].

There are some aspects where the vocal tract differs significantly from an acoustic tube:

- The cross-sectional area varies smoothly throughout the tract.
- The vocal tract bends at back of the tongue. The bend in the vocal tract has little effect on wave-propagation, however it can lower the resonance of the tract by a few percent, due to a decrease in phase velocity [Milenkovic et al., 2010].
- The wall of the vocal tract is not rigid. It will yield to the pressure in the tract at low frequencies and the vibrations of the walls will absorb some of the energy of the acoustic wave.
- The tract branches out to both the oral cavity and nasal cavity if the soft palate is lowered.

In [Sundberg, 1987] it is established that singers learn use their vocal tract to emphasize certain harmonics of the source signal by tuning the frequency of the formants. The singer relies on his/her acoustic feedback to tune both the formants and pitch of the phonation. In [Henrich et al., 2011] it was shown that sopranos tune the harmonics of the voice signal to the higher formants, but the strategy is also used between the first formant and the fundamental, where the fundamental generally was below the first formant for the test subjects in [Henrich et al., 2011].

The air in the vocal tract is heated up to body temperature and is more humid. The speed of sound and the density of air used is: $c = 350$ m/s $\rho_a = 1.13$ kg/m^3[Story and Titze, 1995]

1.3 Preliminary analysis of Complete Vocal Technique

In this section, the concepts taught in CVT will be analysed in the light of the anatomy and physiology.

CVT clearly distinguish between different types of phonation through the modes described in the introduction to this project. This analysis should give a clearer picture of what characterises these four modes of phonation. The CVT textbook [Sadolin, 2012] is used in this analysis, along with data from unpublished studies [Sadolin and McGlashan, 2007], which includes:

- A series of video endoscopy recordings showing the vocal folds and the epilarynx tube during phonation of the CVT modes.
- Data from electroglottographic (EGG) recordings of phonation of the modes.

EGG is a non-invasive technique to estimate the vocal fold contact area. The electric impedance is recorded between two electrodes mounted on each side of the neck at the larynx [Baken, 1992]. The results of the study is a general description of the vocal fold contact area for each mode, using EGG estimations, see figure 1.12.

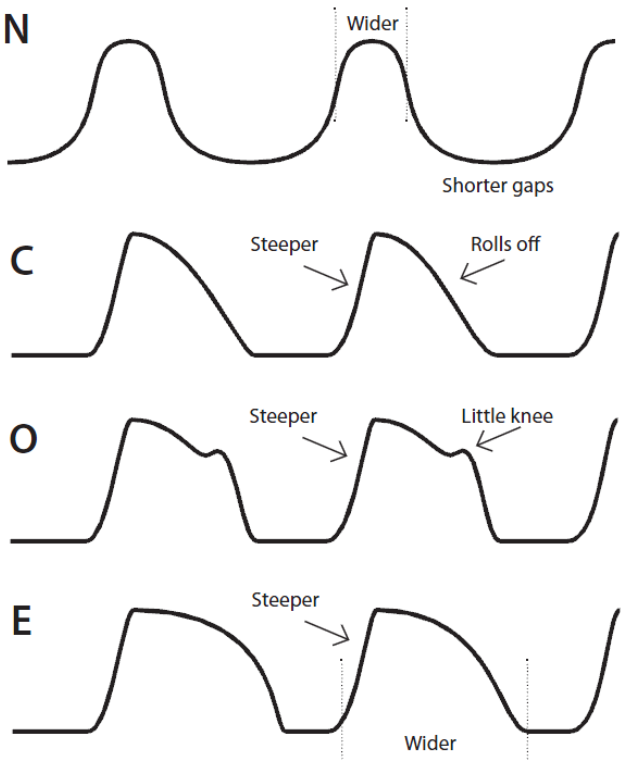


Figure 1.12: Generalized waveforms of the vocal fold contact area of the four modes (N=Neutral, C=Curbing, O=Overdrive and E=Edge) [Sadolin and McGlashan, 2007].

Note that the upwards direction on the y-axis in these plots is the positive estimated contact area, which is not inverse to glottal flow or glottal opening. The waveforms can be used to diagnose singers with problems, but any comparison with physical data should be taken with precautions. The data can be used for an estimation of the duration of the open and closed phase, which is usually described via the open quotient O_q - a factor between 0 and 1 describing the duration of the open phase in relation to the pulse length. See table 1.1.

Mode	Neutral	Curbing	Overdrive	Edge
Open Quotient	.64	.49	.53	.54

Table 1.1: Guidelines for open quotient [Sadolin and McGlashan, 2007].

The derivative of the EGG signal (not shown here) can give an estimation of how fast the vocal folds contact area increases in the closing phase, thus giving clues of the abruptness of the closure [Henrich et al., 2004].

Neutral

Neutral is the softest sounding of the modes and is probably the mode that comes closest to the normal phonation, as described in the physiology section. It is possible to sing in this mode throughout the entire vocal range and in all vowel. However, the loudness of the phonation is quiet in the lowest range, but increases as the pitch gets into the higher ranges and can get loud in the highest ranges. The description of loudness is done according to how it is described in [Sadolin, 2012], from quiet to very loud. In Neutral, a breathy character can be added to the sound, which reduces loudness over the whole range, a constant gap between the vocal folds can be observed on the endoscopy recordings when adding air to Neutral.

The endoscopy recordings show that the epilarynx tube is wide, which would suggest a less efficient coupling between the vocal tract and the vocal folds [Titze, 2008]. This could also explain why it is possible to sing all vowels in this mode and why it is possible to sing in this mode throughout the entire vocal range. The wide selection of vowels also makes it possible to apply different formant tuning strategies. The waveform shows that there is less contact than the other modes, suggesting that the adduction is not as tight as in the metallic modes.

Curbing

Singing in Curbing has a crying/wailing character and can produce a bit louder phonation than Neutral. It is restricted in the use of vowels, where the vowels that are easiest to use Curbing in are: O (as in woman), I (as in sit) and UH (as in hungry).

When learning the technique of Curbing it is explained that one should establish a hold prior to phonation - like holding ones breath. This points toward a tight adduction of the vocal folds. The endoscopy recordings show that configuration of the aryepiglottal fold and the epiglottis is narrowed a little compared to the Neutral mode, but not as much as the full-metallic modes. The Curbing sound is sometimes produced with an opening to the nasal cavity, but it is not a requirement to this mode. The waveform of Curbing and all the metallic modes show a steep closing phase, which has been related to an increase in the strength of the overtones from the source. Curbing has the shortest O_q, which again suggests a tight adduction.

Overdrive

Overdrive has the characteristics of a shout and can produce very loud phonation. It can strictly only be produced on the vowels EH (as in stay) and OH (as in so) and it has an upper limit of C5(522 Hz) for men and D5 (587 Hz) for women. When learning the Overdrive mode it is recommended that a "bite" is established. The "bite" covers a posture of the jaw and lips that create a large opening, with the edges of the lips pulled back towards the ears, see figure 1.13.

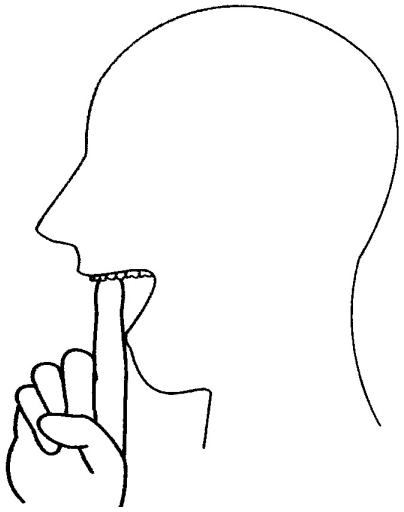


Figure 1.13: Illustration of the "the bite", [Sadolin, 2012].

The bite possibly shortens the effective length of the vocal tract. If a uniform tube is shortened, its resonances move upwards and if the first resonance of the vocal tract is raised, the range of inertive load is widened. The fact that Overdrive can only can be produced up to a phonation frequency of about 587 Hz (D5) leads towards a hypothesis that the first formant is related to the length of the vocal tract in this mode and cannot be tuned upwards.

The endoscopy recordings show that the epilarynx tube is narrowed, forming a tube down towards the vocal folds. The small cross-section above the vocal folds enhances the coupling, thus making an inertive load even more beneficial and a compliant load more destructive for oscillation.

The open quotient points towards a tight adduction of the vocal folds, since the O_q almost is as low as in Curbing. Also the fact that the phonation cannot be obtained in Overdrive at low loudness, points towards tight adduction and perhaps a high subglottal pressure to separate the vocal folds. The waveform in figure 1.12 show a sharp closing phase which would account for the "metallic" in Overdrive, given the amount of overtones.

Edge

Edge sounds more shrill and snarling than Overdrive, but can also produce very loud phonation. However there is no upper restriction of the pitch in this mode. The vowel selection is also wider, as it can be produced at I (as in sit), EH (as in stay), A (as in father) and OE (as in herb).

The concept of twang is known in "popular music" terminology [Sundberg and Thalen, 2010]. In CVT, twang is added to the sound by "twanging" the epiglottic funnel, which refers to a narrowing of the epilarynx tube. CVT describes that the Edge mode can be established by twanging the epiglottic funnel and singing loudly. The Edge mode is described as not giving the "tone" too much space in the back of the mouth, where the tongue held up forming a small opening.

In Edge the open quotient also points towards a tighter adduction of the vocal folds and alse here perhaps a higher subglottal pressure to separate the vocal fold. The waveform in figure 1.12 shows a sharp closing phase. The laryngoscopy of Edge reveals the narrowest epilarynx tube of the modes, which should intensify the coupling even more than in Overdrive.

Summary

The following characteristica about the modes can now be hypothesised.

- Neutral resembles normal phonation, which is loosely adducted and exerts small effects of vocal tract coupling.
- The vocal folds have to be tightly adducted to produce Curbing. The epilarynx tube is narrowed.
- The vocal folds have to be adducted tightly for Overdrive to be produced and the subglottal pressure has to be high. The upper frequency limit may be dictated by a lack of tuning possibility for the first formant. The loudness of the phonation is

due to a strong coupling between the vocal folds and the vocal tract, as the epilarynx tube is narrow.

- In Edge, the vocal folds also have to be tightly adducted and the epilarynx tube is even narrower than overdrive, which should intensify the coupling even more. The loudness of the phonation is due to the strong coupling between the vocal folds and the vocal tract.

It is believed that tuning the first formant may have an effect on the phonation of the modes. A part of the timbre change between the modes can be attributed to the change in vowels.

1.4 The purpose of the further analysis

The following analysis of the vibro-acoustics of the singing voice will focus on the analysis and modelling of some of the characteristics of the four modes of Complete Vocal Technique.

The analysis requires a model to be formulated that can simulate the vibro-acoustics of the phonation of these modes. This model must be capable of modelling the parameters that have been hypothesised as key factors for phonation of the modes. To summarize, the model should be able to:

- Simulate vibration of the vocal folds with different degrees of adduction.
- Simulate the effects of the coupling between the vocal folds and the vocal tract.
- Simulate phonation at the given pitch

A study of the vocal tract in the given modes will be conducted, to which data will be obtained through MRI imaging of phonation in all the modes.

Audio recordings will be made of all the modes, which will be analysed to be compared with modelled data.

CHAPTER 2

Modelling the system

The tools for modelling the vocal folds and the vocal tract are explored before the analysis can begin. In the following chapter the concepts that has been established so far is interpreted for modelling and a model will be chosen and described. The model include:

- A two-mass model of the vocal folds
- A simple description of the aerodynamics in the glottis and the distribution of aerodynamic pressure on the vocal folds
- A frequency-domain model of the vocal tract, including losses.
- A time-domain model of the vocal tract, including losses.

Vocal fold modelling

Generally there are two approaches to modelling the vocal folds: The lumped parameter model and the continuous model. The continuous model uses FEM or other numerical analysis tool for boundary value problems to model the dynamics of the mechanics and fluid dynamics. It requires very detailed knowledge of the geometrical and material properties of the vocal folds. Due to time considerations and simplicity such a model is not considered for this project.

The lumped parameter models describe the vocal fold vibration using simple damped mechanical oscillators, a calculation of the flow through the glottis and a calculation of the distribution of the aerodynamic forces on the mechanical oscillators. Many of the lumped parameter models were designed for the purpose of speech synthesis, but has shown to be a beneficial tools in physiological research. To find the most suitable model for this project, a study of the established models was done, which can be found in appendix A.

Most lumped parameter models use the symmetry between the two vocal folds so calculations only are carried out for one vocal fold. It is also very common to simplify the

opening area of the glottis as a rectangular duct and the vibration in the plane along the vocal tract is also often neglected [Titze and Strong, 1975].

The classical vocal fold model consists of two coupled masses, that can model the first string mode of the vocal folds and the shearing mode of the mucosal wave. Many other designs have been proposed to model the movement of the body/cover structure. The simplicity of the lumped parameter models often come with the cost of physiologically relevant parameters, even though the parameters for these models have been found through experiments with exised larynges and in-vivo experiments [Titze and Alipour, 2006]. This makes the requirements for validating measurements greater. It is a priority in this project to keep the modelling as simple as possible as to limit the number of adjustable parameters and to simplify the calculations.

The models in general all assume that the flow is quasi-steady, which means that static flow equations can be used to describe the flow even though the system has dynamic boundaries. The flow is updated in each time-step of modelling, taking the instantaneous geometry of the vocal folds into account. Often all time dependency in the flow, such as viscous and inertive effects are neglected. The approximation is especially crude in the closing phase of the vocal folds, where the viscous effects in the flow is large. However this necessary for the simplicity of the model. [Pelorson et al., 1995].

Chosen vocal folds model

A simple two-mass model is chosen to model the vocal folds in this project. The model combines aspects from [Birkholz et al., 2011], Steinecke and Herzel [1995] and [Story and Titze, 1995].

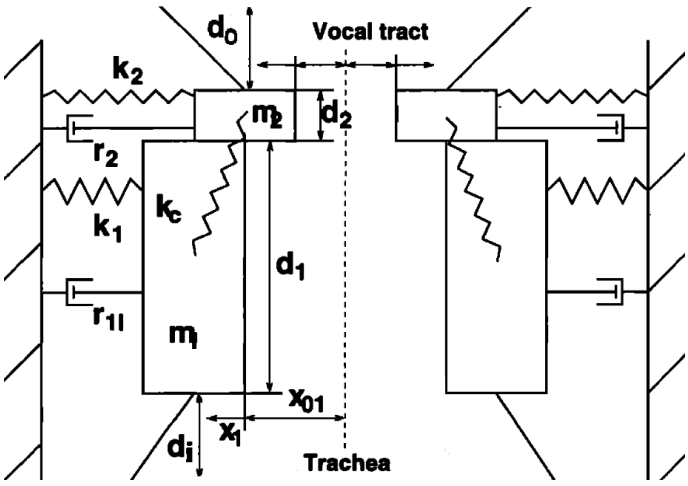


Figure 2.1: Illustration of the mechanical system of the two-mass model.

Mechanical system

The mechanical system is illustrated in figure 2.1, where the symmetric vocal folds are shown. An analogy for the normal movement of the system is that of a parent running with a child. The horizontal movement of the child will lag behind the parent. It is an important part of this model that the upper mass is smaller than the lower mass. The movement of the masses is described by their equations of motion:

$$F_1 = m_1\ddot{x_1} = F_{k1} - F_{kc} + F_{kCol1} + F_{r1} + F_{e1}, \tag{2.1}$$

$$F_2 = m_2\ddot{x_2} = F_{k2} + F_{kc} + F_{kCol2} + F_{r2} + F_{e2}, \tag{2.2}$$

where F_{k1}, F_{k2} are spring forces, F_{kc} is the coupling spring force, F_{r1} and F_{r2} are damping forces, F_{kCol1} and F_{kCol2} are the spring forces during collision and F_{e1} and F_{e2} are the external forces from the aerodynamic pressures. The spring forces are given by the following equations:

$$F_{k1} = -k_1(x_1 - x_{10}), \quad F_{k2} = -k_1(x_2 - x_{20}), \tag{2.3}$$

$$F_{kc} = -k_c\left[(x_1 - x_{10}) - (x_2 - x_{20})\right]. \tag{2.4}$$

The damping forces are given by the following:

$$F_{r1} = -r_1\dot{x_1}, \quad F_{r2} = -r_2\dot{x_2}, \quad \text{where} \quad r_1 = 2\zeta_1\sqrt{m_1k_1}, \quad r_2 = 2\zeta_2\sqrt{m_2k_2}. \tag{2.5}$$

When the masses cross the mid-line, the collision is triggered through the boolean function $\Theta(x)$

$$\Theta(x) = 0, \quad x > 0 \tag{2.6}$$

$$\Theta(x) = 1, \quad x \leq 0 \tag{2.7}$$

The collision is modelled by adding stiffness and damping to a mass when collision is active.

$$F_{col1} = -\Theta(x_1)\left(h_1x_1 + 2\zeta_{h1}\sqrt{m_1h_1}\dot{x_1}\right), \tag{2.8}$$

$$F_{col2} = -\Theta(x_2)\left(h_2x_2 + 2\zeta_{h2}\sqrt{m_2h_2}\dot{x_2}\right). \tag{2.9}$$

The open areas between the masses in the glottis is described as:

$$a_1 = 2l_gx_1, \qquad a_2 = 2l_gx_2 \tag{2.10}$$

The parameters used for the mechanical system are shown in table 2.1. They are inspired by [Birkholz et al., 2011], but the collision damping have been increased and outlet of the glottis have been removed to make the model more stable. The phonation frequency of the

model is adjusted using the q parameter, when $q = 1$ the phonation frequency is approx. 130 Hz. The relationship between the masses and stiffnesses will simply determine the resonance of the system by:

$$\omega_0 = \sqrt{\frac{k_1 + k_2}{m_1 + m_2}}. \tag{2.11}$$

However, the non-linearities and the damping of the system will alter the phonation frequency. This parameter therefore have be be adjusted to obtain an exact pitch.

The equations of motion is solved using a simple finite difference solution. The choice of this type of time-domain solver is done because the time-domain model of the vocal folds has specific requirements to the computation of the sampling. The choice is justified by a very high sampling frequency compared to the vibration frequency of the mechanical system. The non-linearities may cause instabilities when using this type of solver, but it was not experienced.

Description	Parameter	Value
Mass 1	m_1	0.125/q g
Mass 2	m_2	0.025/q g
Stiffness 1	k_1	80·q N/m
Stiffness 2	k_2	8 ·q N/m
Coupling stiffness	k_c	25·q^2 N/m
Thickness 1	d_1	2.5/$\sqrt{q}$ mm
Thickness 2	d_2	0.5/$\sqrt{q}$ mm
Inlet thickness	d_i	4 mm
Glottis length	l_g	13 ·$\sqrt{q}$ mm
Damping 1	ζ_1	0.1
Damping 2	ζ_2	0.6
Additional collision damping 1	ζ_{h1}	0.9
Additional collision damping 2	ζ_{h2}	0.43

Table 2.1: Parameters for the two-mass model, partly from [Birkholz et al., 2011]

Aerodynamics

A simple model of the aerodynamics is implemented mainly from [Story and Titze, 1995]. The pressure on the lower mass is modelled as changing between Bernoulli and jet flow in the converging and diverging glottis, as described in chapter 1.2, where switch between the two types of flow happens abruptly in this model. The following expression describes

the average pressure over the lower mass:

$$p_{1,Bernoulli} = P_s - (P_s - P_e)\left(\frac{\min(a_1, a_2)}{a_1}\right)^2. \tag{2.12}$$

Choosing the minimum area of the two masses makes sure that the lower mass is only subjected to p_e when the glottis is diverging. When $a_2 = 0$ the lower mass will be subjected only to p_s. When the glottis is converging, the average pressure drop ratio over the mass is described by $(a_2/a_1)^2$ [Story and Titze, 1995]. The upper mass is always subjected to jet flow, hence the average pressure is:

$$p_{2,jet} = P_e. \tag{2.13}$$

When collision is detected on either of the masses, the aerodynamic pressure at it is zero. Using [Birkholz et al., 2011], the pressure on the inlet of the glottis is taken into account in the calculation of the external forces to the mechanical system F_{e1} and F_{e2}. Without the inclusion of the inlet it is not possible to simulate negative adduction. The calculation of the forces are simply done by multiplying the pressure over the area of the mass, with an addition of a calculation of the hinge moment for the inlet pressures:

$$F_{e1} = p_1 t_1 l_g + 0.25(p_s + p_1) d_i l_g, \qquad F_{e2} = p_2 t_2 l_g. \tag{2.14}$$

The flow through the glottis is calculated through the closed form solution derived in [Titze, 1984]. The formulation generally uses the Bernoulli's equation to describe the flow through the glottis given the minimum area. A coefficient k_t is implemented to compensate for the viscous losses. The pressure wave in the vocal tract travelling towards the glottis and an expression for the equivalent area between the beginning of sub-glottal tract and vocal tract is included in the pressure drop.

$$u_g = \frac{a_g c}{k_t}\left[-\frac{a_g}{A^*} \pm \sqrt{\left(\frac{a_g}{A^*}\right)^2 + \frac{4k_t}{c^2\rho}(p_s/2 - p_e^-)}\right], \tag{2.15}$$

where the positive sign is used when the term inside the square root is positive. a_g is the minimum area of the glottis and the equivalent area A^* is given by:

$$A^* = \frac{A_s A_i}{A_s + A_i}, \tag{2.16}$$

where A_s and A_i is the area of the input and output of the glottis, respectively. p_s is the static lung pressure. In this project the pressure fluctuations in the sub-glottal tract have been neglected.

2.1 Vocal tract models

Two models of the propagation of acoustic waves in the vocal tract is implemented in this project. A frequency-domain model and a time-domain model. The frequency-domain model is a simpler model, but it is very convenient to have the model in the time-domain when coupling to the vocal folds model.

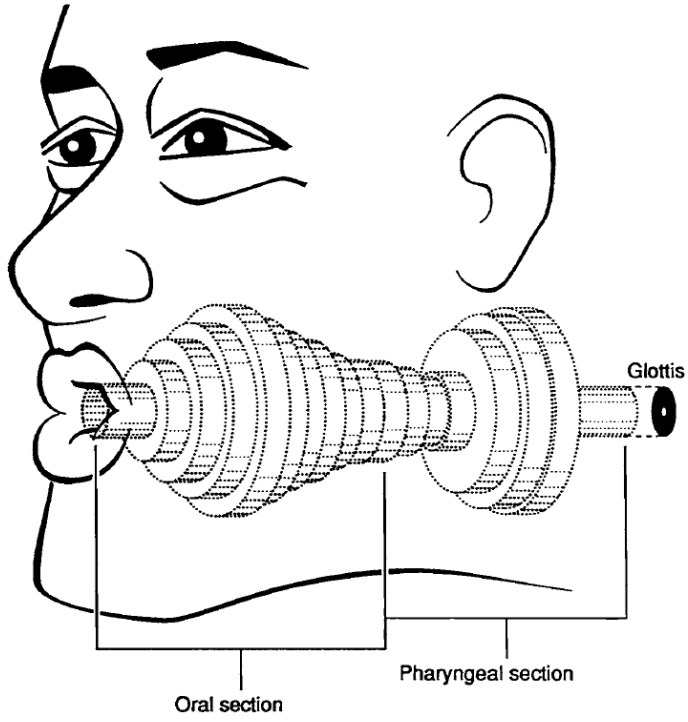


Figure 2.2: Illustration of the vocal tract models [Story, 1995].

Both these models divide an area-function of the vocal tract up into a number of coupled, uniformly long, circular tube sections of different cross-sectionas illustrated in figure 2.2 - 44 section are used in this project. This method can be used given the approximation that the sound field in the tube system is one-dimensional. For tubes of a uniform circular cross-section it is a good approximation well below where standing waves appear in the cross-section of the tube. This will happen where the largest distance is comparable to half the wavelength. Given the later analysis the largest cross-section is about 3 cm's, which would produce standing waves at approximately at 5.5 kHz. Therefore a frequency limit of 4 kHz is used throughout the modelling.

The non-uniformity of the vocal tract is very crudely approximated with tubes of uniform cross-section, especially at higher frequencies, but it is a very convenient way of modelling the problem. In this project numerical modelling using Comsol will be used to verify the models.

The acoustic characteristics of a tube is often described by its impedance Z. The input impedance Z_i of a tube is the complex response of the sound pressure p at the input of the tube, to a given volume velocity input u. The transfer impedance describes the complex pressure at the output of the tube, to the same given input.

An acoustic load is characterized as inertive, if the phase of the acoustic pressure p is 90° ahead of the phase of the volume velocity u. For a compliant load the phase is 90° behind instead. The concept can be described by the analogy of pushing the column of air in a tube, when the load is a pure mass load, the air moves as one structure through the tube. If the load is compliant it is compressed when pushed. These concept covers how the energy of a acoustic wave is stored. The inertive load is storing the energy of the oscillation in kinetic energy and the compliant load is is storing it in potential energy. The complex impedance to an acoustical system characterises whether the load is inertive, compliant or resistive. If the reactance (imaginary part) is positive, then the load of the system is characterised as inertive and if the reactance is is negative the load is compliant. If the impedance is purely real the load is resistive. [Titze, 1988].

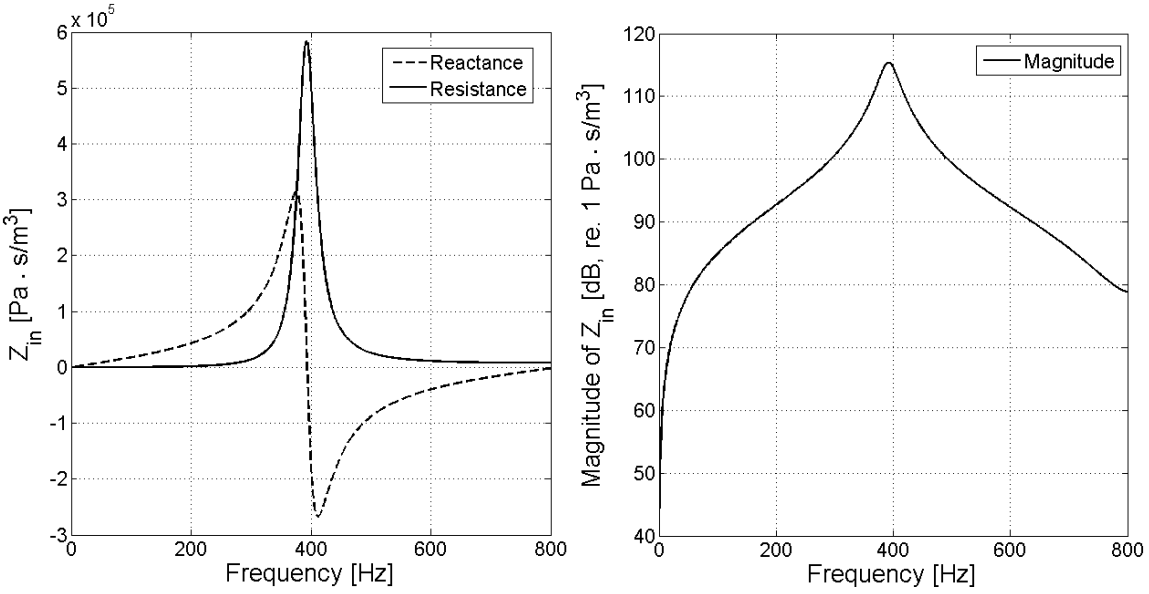


Figure 2.3: Illustration of the input impedance of a tube terminated by a radiation baffled impedance(length=17 cm, cross-section=100 cm^2).

Figure 2.3 shows the relationship between the reactance and the resistance of a uniform tube, terminated by a radiation impedance,. The load of the input impedance to this tube is inertive up to where the resistance start to take over at the first resonance of the tube. The load then changes to compliant in frequencies above the resonance.

Frequency-domain model

The frequency domain model uses the calculation of the impedance of coupled pipes to model the acoustics in the vocal tract. The input impedance of one tube section is described by:

$$Z_i = \frac{AZ_t + B}{CZ_t + D}, \tag{2.17}$$

where Z_t is the termination impedance, and A, B, C and D are transmission coefficients:

$$A = \cos kl, \quad B = j\frac{\rho c}{S}\sin kl, \quad C = j\frac{S}{\rho c}\sin kl, \quad D = \cos kl, \tag{2.18}$$

This description is derived from the one dimensional wave-equation in a tube with rigid walls [Jacbosen, 2011]. The system of coupled tube sections is the calculated by letting the termination impedance of one tube, be the input impedance of the next. At the end of the series of tube sections the termination impedance depends on the modelled situation, which could be a radiation impedance or a blocking of the tube with an infinite impedance. The radiation impedance of the mouth is usually approximated by the radiation impedance of a piston in a baffe [Story et al., 2000].

$$Z_r = \frac{j\omega RL}{R + j\omega^2 L}, \quad \text{where:} \quad R = \frac{128\frac{\rho c}{S}}{9\pi^2} \quad \text{and} \quad L = \frac{8\sqrt{\frac{A_{mouth}}{\pi}}\frac{\rho c}{S}}{3\pi c}. \tag{2.19}$$

The losses of the vocal tract is implemented as in [Story et al., 2000], where the frequency-domain model is modified. It builds on a solution of the one-dimensional wave-equation in a tube with a simplification of the yielding boundaries and losses to calculate the impedances of each tube section [Sondhi, 1974]. The effect of the yielding wall is a slight increase of the first resonance frequency [Story, 1995].

The frequency-domain model can be used to calculate the radiated pressure on-axis at a distance r. The transfer function between the flow through the glottis and the volume velocity at the lips is first calculated:

$$q_{out} = \frac{q_{in}}{CZ_{\text{term}} + D}, \tag{2.20}$$

where C and D are the coefficients in the transmission matrix for the frequency domain model, calculated to the end tube section. Z_{term} is the radiation impedance. This method is assuming that the glottal source impedance is much larger than the radiation impedance. The on axis pressure at a distance r is approximated by the radiation of a point monopole with the volume velocity q_{out}:

$$|p(r)| = \frac{\rho\omega|q_{out}|}{2\pi r}. \tag{2.21}$$

The transfer function from glottis to the receiving point is calculated by combining these two equations.

2.1.1 Time-domain model

A wave reflection model (WR model) is used for modelling the vocal tract in the time-domain [Story, 1995]. The model relies on the solution of the one-dimensional wave equation in the time domain. As the name implies, it works by calculating how the pressure waves are reflected at the junctions between the tubes.

The model is synchronized so that over one time step a pressure wave will propagate exactly the distance to the next tube section, in this way discretizing the wave propagaton. See figure 2.4 for an illustration. This method specifies that the model is restricted to having a fixed sampling frequency of:

$$F_s = \frac{Nc}{l} \approx 88 \text{ kHz}, \tag{2.22}$$

where N is the number of tube sections and l is the length of the vocal tract.

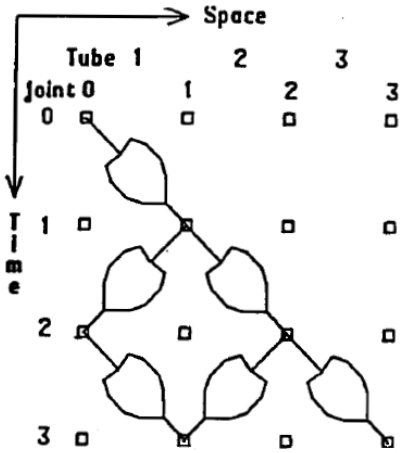


Figure 2.4: Illustration of the wave propagation in the WR model [Story, 1995].

When a pressure wave is injected into the system at the glottis it propagates towards the mouth. If it meets a change in cross-sectional area the pressure and volume velocity relationship changes given the impedance of the tube. Part of the wave will continue to propagate towards the mouth and part of it will propagate back towards the glottis.

The coefficients describing the reflections is found by solving the one-dimensional wave equation in the junction between two tube section. The reflection coefficients $r_{k,1}$ and $r_{k,2}$ describe the reflected pressure when the pressure wave comes from either side:

$$r_{k,1} = \frac{A_k - A_{k+1}}{A_k + A_{k+1}}, \qquad r_{k,2} = \frac{A_{k+1} - A_k}{A_k + A_{k+1}}. \tag{2.23}$$

The output of a tube junction is then described by the following scattering equations:

$$B_k = F_k r_{k,1} + B_{k+1}(1 + r_{k,2}), \qquad F_{k+1} = F_k(1 + k, 1) + B_{k+1} r_{k,2} \tag{2.24}$$

In the first section of the vocal tract the flow through the glottis is injected by the following relationship:

$$F_1 = u\frac{\rho c}{A_1} \tag{2.25}$$

The reflection at the glottis is implemented by adding a tube section of zero area at A_0. Full reflection is kept throughout the vocal fold oscillation to avoid unnecessary discontinuities from the non-linearities of the glottis.

When a pressure wave hits the end of the last section of the vocal tract, part of the wave is reflected by the radiation of the lips. As in the frequency-domain model, this is described by the radiation impedance of a piston in a baffle. The reflection coefficient of the radiation impedance is given by the following frequency dependent reflection coefficient:

$$r_m = \frac{Z_{rad}(j\omega) - Z_{mouth}}{Z_{rad}(j\omega) + Z_{mouth}}. \tag{2.26}$$

To implement the frequency dependency in the time-domain it is necessary to use a bilinear transformation. The implementation of this into the scattering equation is described in [Story, 1995].

The losses of the vocal tract are also modelled in the time-domain model. The losses have a stabilizing effect the system when it is coupled to the model of the vocal folds. Without the losses, the model very easily becomes unstable whenever a resonance is excited.

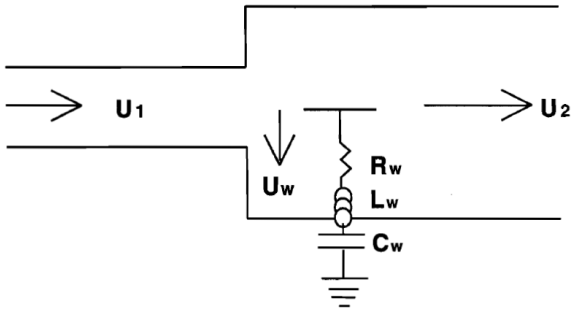


Figure 2.5: Illustration of the yielding wall model.

The implementing of the effects of the yielding wall and the losses are described in detail in [Story, 1995]. The yielding wall is modelled in each tube section by part of the volume velocity exciting a mass, spring and damper system modelling the wall, as illustrated in figure 2.5.

The system is implemented in the time domain by a central difference algorithm. The implementation of the losses and reflections caused by the yielding wall are incorporated into the scattering function.

To match the losses with the ones implemented in the frequency-domain model, the parameters are found in [Man and Schroeter, 1987]. Additional losses were added to

compensate for the viscous losses and for stability in the time-domain model. These were simply approximated by removing a small percentage pressure in the model in each timestep [Liljencrants, 1985].

Validation of the vocal tract model

The input impedance of an area-function is calculated using both models. To find the input impedance of the WR model, an impulse of particle velocity $u = 1$ is injected at the glottis at $t = 1$ and the pressure and volume velocity response is recorded. The input impedance is calculated from the recorded pressure over the volume velocity in the frequency domain using a DFT algorithm. The results are shown in figure 2.6

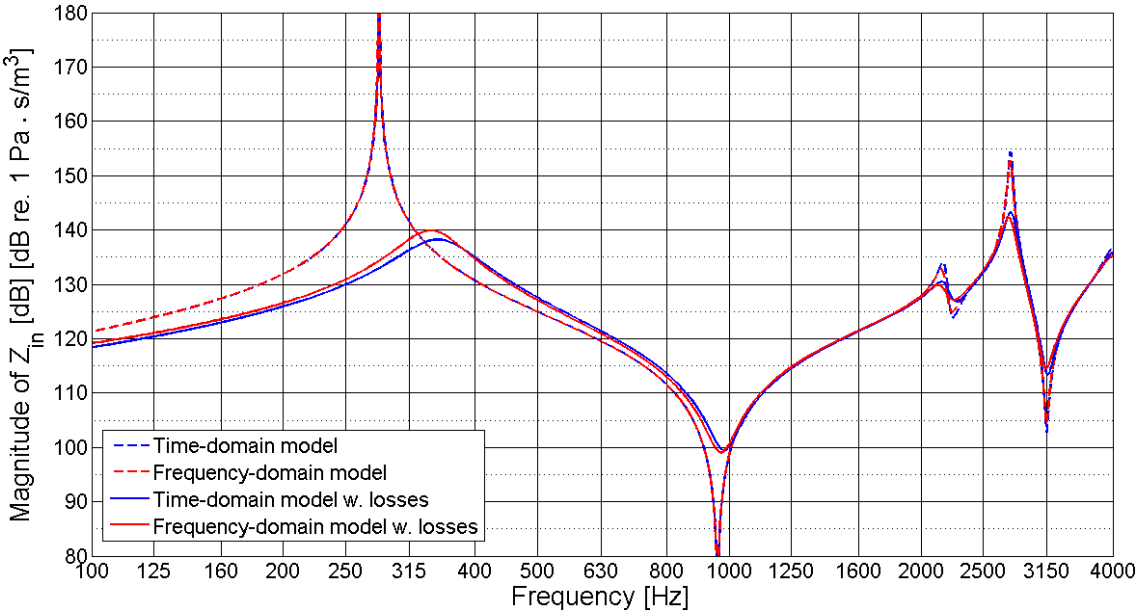


Figure 2.6: Verification of the WR model. Comparison between input impedance of the extracted area function for Neutral, using both models of the vocal tract w. and without losses.

The results show that there are differences between the results of the input impedances of two models. The damping of the resonances and the yeilding wall are not matched exactly, but the method of implementing them also differ between the models. The desired effect is modelled in the WR model and this will be used in the further study.

CHAPTER 3

Acquiring vocal tract data

The preliminary analysis pointed out that the vocal tract may have a significant effect on the phonation of the CVT modes. The vocal tract will shape the signal from the glottal source into the vowels that are used in the modes, but the coupling between the acoustics of the vocal tract and the source is suspected to play a role in these types of phonation.

To be able to implement data about the vocal tract in the chosen model it is required to have a one-dimensional area-function describing the cross-sectional changes through the vocal tract. Geometrical data of the vocal tract has been acquired numerous times before, using various medical imaging techniques, f.ex. in [Clement et al., 2007]. However, since no studies have been done on the vocal tract shapes when phonating CVT modes it is necessary to perform new measurements. The possibility of using an ultrasound technique was investigated prior to the MRI, but it was not well-suited for this application. In collaboration with Lars Hanson at DRCMR and Hvidovre Hospital it was possible perform the MRI scans for this project.

Data on the vocal tract should be acquired on the testsubject while phonating in the four modes, plus Neutral with flageolet (was not analysed in this project, see [Sadolin, 2012]). It was decided to do a study using just one pitch. In this case 261 Hz (C4) throughout all recordings.

This chapter includes:

- A description of the procedure for the MRI scans.
- Extraction of the area-function from the airway geometry in the MRI recordings.

3.1 Acquiring geometrical data through MRI

The basics of Magnetic Resonance Imaging can be studied in [Hanson, 2009]. The method applies extremely big magnetic fields to make the water molecules in the body resonate, emitting radio frequency waves. The image acquired using MRI is therefore mainly a three-dimensional map of the concentration of water molecules in the body.

In this study, four scan sessions were performed in total, where two of them were pilot sessions, testing how the recordings should proceed to acquire the best results. Those were done on the author and on supervisor Finn Agerkvist. Another pilot session was performed with the singer before the actual session on the singer took place.

Prior to the pilots the following points were set up as requirements to the procedure:

- The scanning cannot take longer than the singer can phonate steadily. The singer may be able to phonate for up to 30 seconds, but the posture of the body changes when letting out the air, which can blur/skew the recordings.
- The boundaries between the airway and the tissue should be relatively sharp.
- The position of the vibrating vocal folds is visible for starting point for the geometry.

Through the experience with the pilot scannings it was decided to use a type of MRI procedure called FLASH, which lightly excites the same slice of tissue repeatedly with intervals of a few ms. The contrast in these recording will show a combination of the amount of water in the tissue and information about the texture of the tissue. Recordings made with this method will be in stacks of slices with a good in-plane, but a worse resolution out of the slice-plane. The recordings were made with a resolutions of 1 x 1 x 3.6 mm/pixel, where the 3.6 mm is the height of each slice. The contrast of the images made it well suited for showing the boundary between the tissue of the airway and the air, but the slice should be recorded perpendicular to the airway to minimize the effects of the poor out-of-plane resolution.

A lower height of the slices could be obtained by recording the odd slices first and the even slices later in the process. However, they were slightly misaligned due to movement of the singer during phonation. It was therefore chosen to scan in geometric order. The singer also moves, but the recordings will not show a misalignment on every second image, which complicates the analysis part. The error caused by the movement is small and is ignored in the geometrically ordered slices.

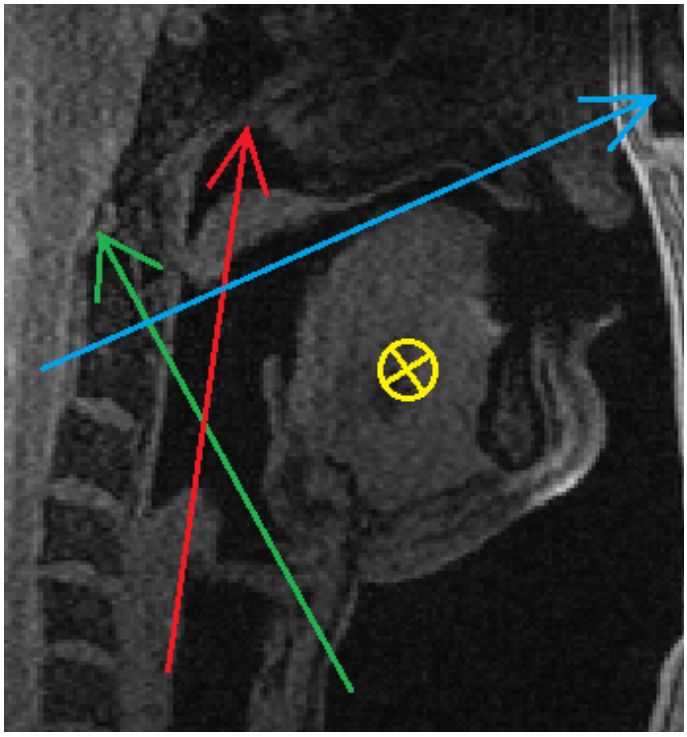


Figure 3.1: The sliceplanes illustrated by their normal vectors. The image is a slice in the saggital plane while phonating in Neutral.

As shown in the figure, four slice planes were recorded, each containing 40 slices. It was decided to use four planes to get a perpendicular slicing of the vocal tract in all positions. The recording of one stack took 11 seconds, so one stack was recorded per phonation to minimize the movement. The compromise of acquiring the data across the different phonations was necessary to acquire all the data. The stacks represented by the the arrows cover the following regions in good resolution:

1. The red arrow covers the pharyngeal cavity
2. The green arrow covers the airway through the larynx.
3. The blue arrow covers the oral cavity
4. The yellow arrow covers all, but is used mainly for the oral cavity.

A drawback to using MRI is that the teeth does not show in the images. However, the black region around the lower front teeth in figure 3.1 is artifacts because the test subject has a metal dental brace glued to the back of his lower front teeth. The artifacts are in a small region of the oral cavity, but it does cover the tip of the tongue and the part of the lower lip. The cavity in the area will be reconstructed from pictures taken of the singers mouth during phonation and from estimation of where the tongue and teeth are located.

It was decided that the order of the recordings was as follows:

- Slice-plane 1 recorded with Overdrive, Edge, Curbing, Neutral and Neutral w. flageolet (all pitch C4)
- Slice-plane 2, same order, same pitch
- Slice-plane 3, same order, same pitch
- Slice-plane 4, same order, same pitch

The reason for this order is that it was easier to continue scanning using the same slice settings. However, it may have resulted in a variation in way the modes were sung, because of the time passed between the same type of phonation.

An expert (Cathrine Sadolin) was present at the scanning session, to verify that all the modes were in fact sung in the correct mode. This was however done over the intercom to the scanner room, where the phonation was partially masked by the scanner noise, especially in Neutral. The phonation was verified in the second before the scanner started and the singer was instructed to hold phonation for a second after the scanner stopped, to leave some time to verify that the phonation had been held in the correct mode and pitch.

The noise of the scanner was not only masking the phonation in the intercom, but also the test subjects inner feedback. Therefore there is an very slight uncertainty that the phonation was completely correct within the modes. The fact that the singer participated in the pilot recording before the real measurement is assessed to have lowered the probability of this happening, since he had gotten somewhat used to being and singing in the environment of the scanner noise.

3.2 Image analysis and extraction of area-functions

It is possible to obtain the area functions by direct analysis of the slices individually. In [Clement et al., 2007] this is done by manually drawing up the edge of the canal from the MRI images and calculating the area of the from this outline. In this study, three dimensional surface meshes were created of the airway and all further analysis was performed on these. The extraction of the area function is performed in MATLAB.

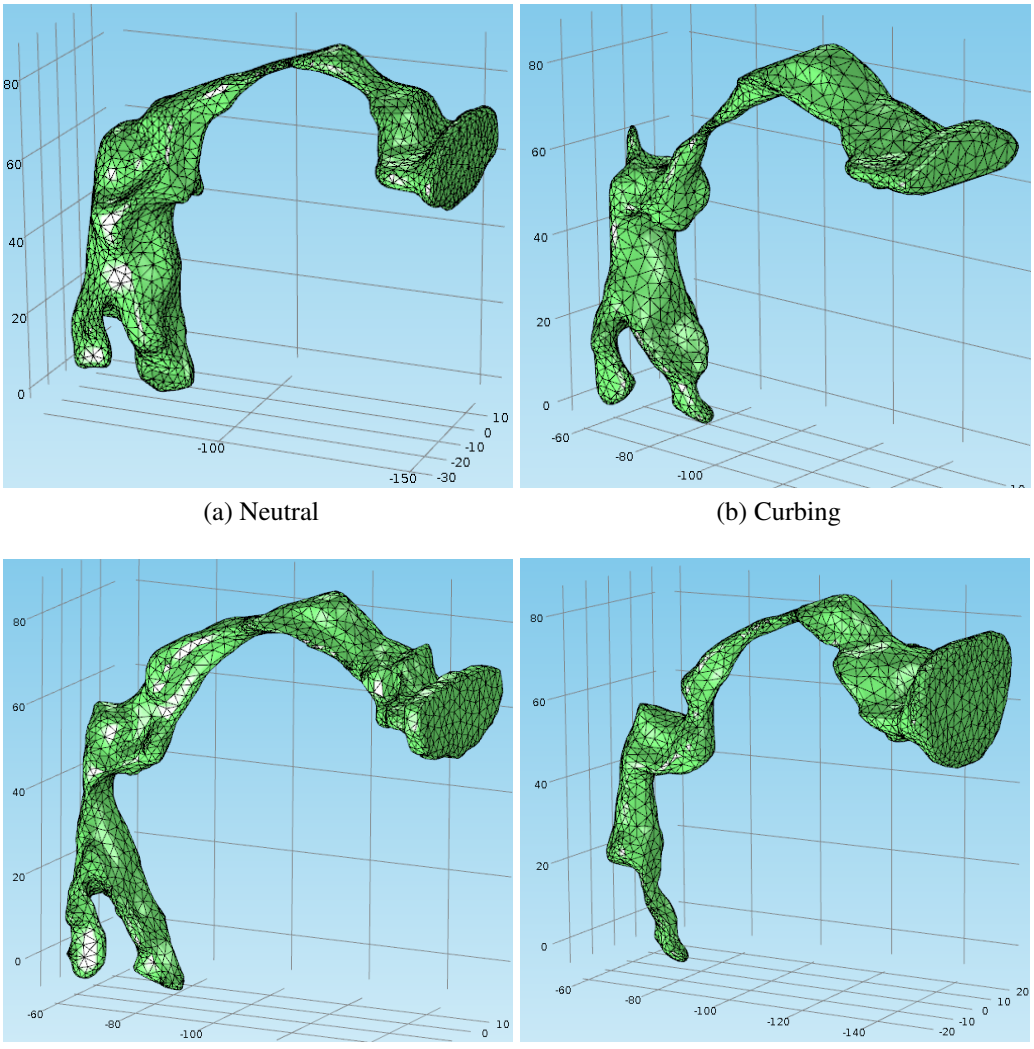
Image analysis

The software ScanIP is used to pre-process the images and segment out the airway. A detailed description of this process is found in appendix B. The points and results of this appendix is summarized here.

It was decided to mainly use the recordings of the slice-planes 2 and 4 in the further analysis. To simplify the analysis a single surface mesh of the entire system was from series no. 4 and only the epilarynx tube was segmented out of no. 2. Series no. 4 was chosen since it displays the whole system in good quality and a comparison of the epilarynx tube is rather consistent with the data from the series no. 2. It is noted that the passage to the nasal cavity is open in the MRI recordings of curbing, which may cause deviations in the following analysis.

Figure 3.2: The mask of a slice in the saggital plane (Neutral mode), showing the segmentation of the airway

Prior to the segmentation, part of the noise was removed with a a gradient anisotropic diffusion noise removal filter. This is an algorithm that removes noise while still preserving the edges in the image [Farsight, 2009]. A three dimensional mask was layered onto the airway using the a combination of the automatic and manual functions in the software, see figure B.3. The figure also shows the results of the reconstruction of the teeth and tongue, which was done on all the segmentations using photos and estimations of the tongue and teeth placement. The remaining reconstructions can be found in the appendix. The mask was the smoothed out using gaussian blur to simplify the created surface mesh. In figure B.7 illustrations of the result of the segmentation is found. The 3D meshes are shown here with an opening to the piriform sinuses, which are removed in the extraction of the area functions.



(a) Neutral

(b) Curbing

(c) Overdrive

(d) Edge

Figure 3.3: The surface meshes of the segmented airway of the four modes

Extraction of the area function

The surface meshes are imported into MATLAB for the extraction of the area functions. The fine details of this processing can be found appendix C. Only the main points and results are described here. The one-dimensional area-functions are created from the areas of the outlines of the vocal tract, which are put in a trail of slices through the meshes.

The vocal tract wall is mapped out by the intersection between the surface mesh and a slice plane. Each slice plane is defined by an origin point on the slice plane and another point that maps out the normal vector of the slice plane. The origin points are placed on a trail through the vocal tract that maps out an approximation of how the plane wave travels. The points that map out the normal vectors are placed so that the normal vector is perpendicular to the vocal tract. This is assumed to create slices through the vocal tract that are perpendicular to the wave propagation direction. The trail of the slices is optimized through the verification of the area function explained the next section. Generally the trail takes the longer way through the vocal tract to compensate for the drop in phase velocity in the bend on the vocal tract. An example of the mapping of the trail through the vocal tract is shown for Overdrive in figure 3.4. Here a long trail and a short trail has been chosen to show the differences.

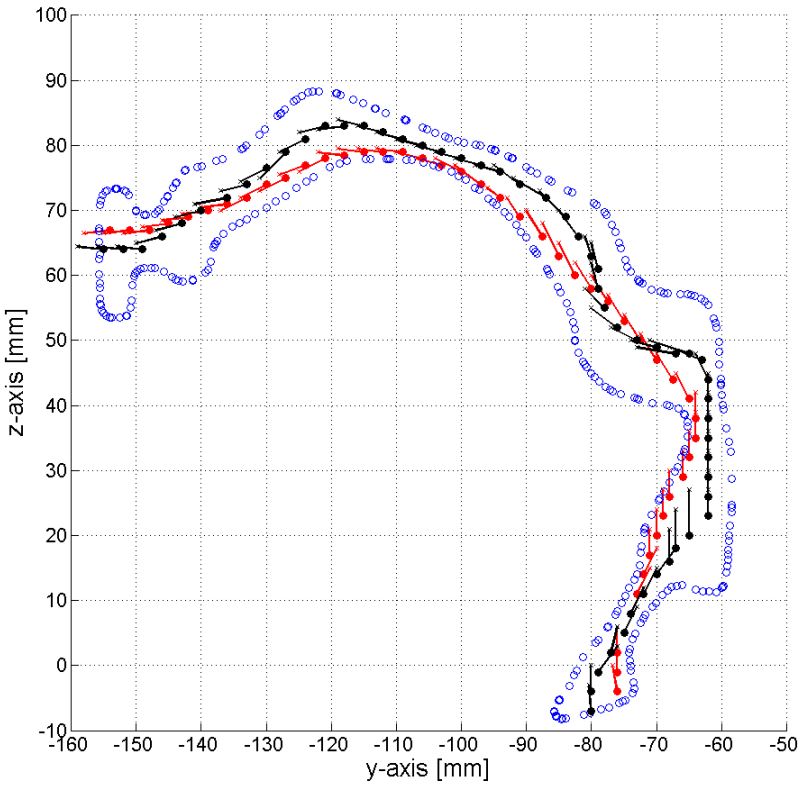


Figure 3.4: Example of two trails of slice selection on a slice in the yz-plane. The red dots and normal vector maps out the short trail and the black dots and normal vector the longer trail.

The distance from the glottis to the slice planes is calculated using the distance between the origin points. The cross-section of the vocal tract at each slice is outlined by the intersection points between the slice plane and the triangles of the surface mesh. The area inside the outline is extracted by connecting the intersection points in a polygon and calculate the area of it. The area function of the epilarynx tubes have been extracted from the results of scan series no. 2 and inserted in the area functions. Also the open area of the mouth and teeth is altered to fit the images, see appendix B for the analysis of these.

An experiment was made to test the sensitivity of the segmentation process. It is justified to add 0.5 mm the equivalent radius of every circular tube element in the area-function. The resolution of the segmentation was 1 mm and the segmentation was usually drawn up on the inside of the airway. This is done both to show the sensitivity and to improve the predictions of the model. The result is shown along with the results of the original area-function in figure 3.5.

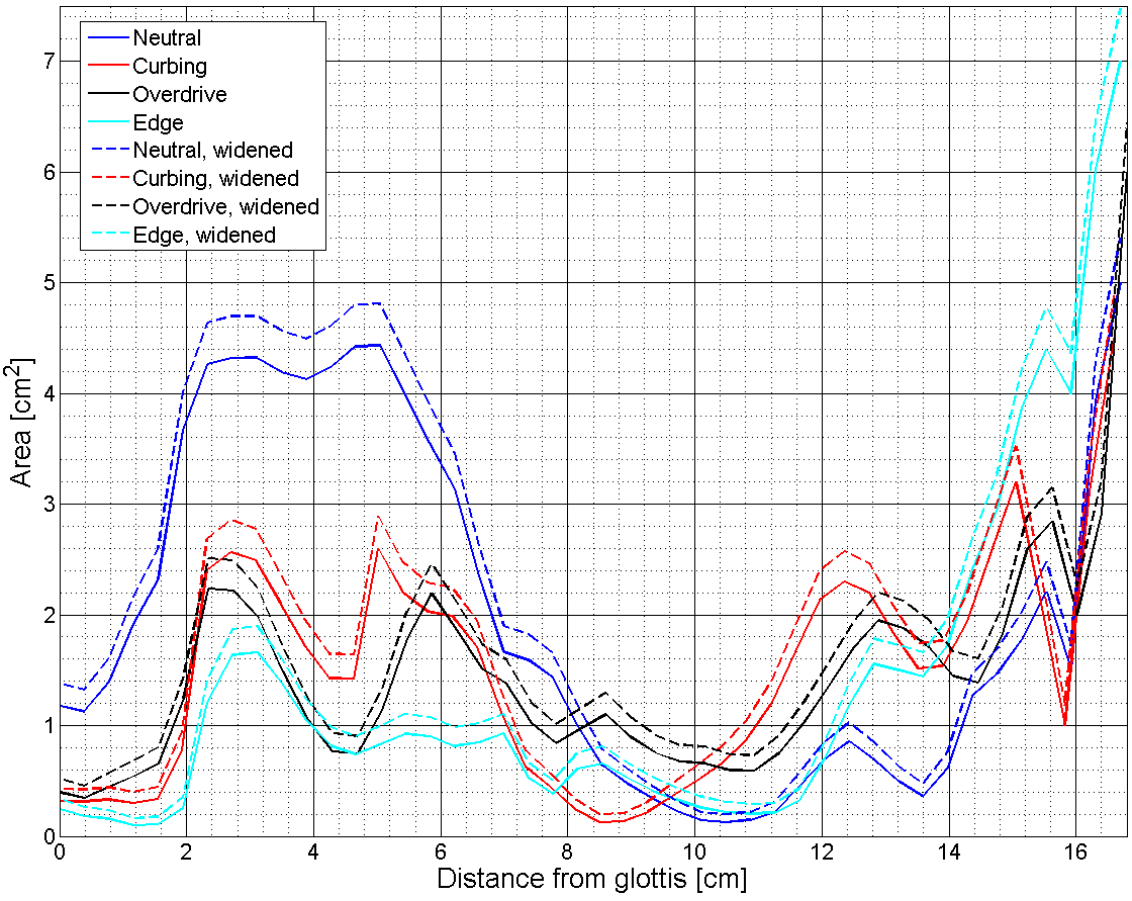


Figure 3.5: Area functions extracted from the surface meshes of the 4 modes.

The transfer function is simulated for both the widened area-function and for the original area-function using the frequency-domain model, see the results in figure 3.6.

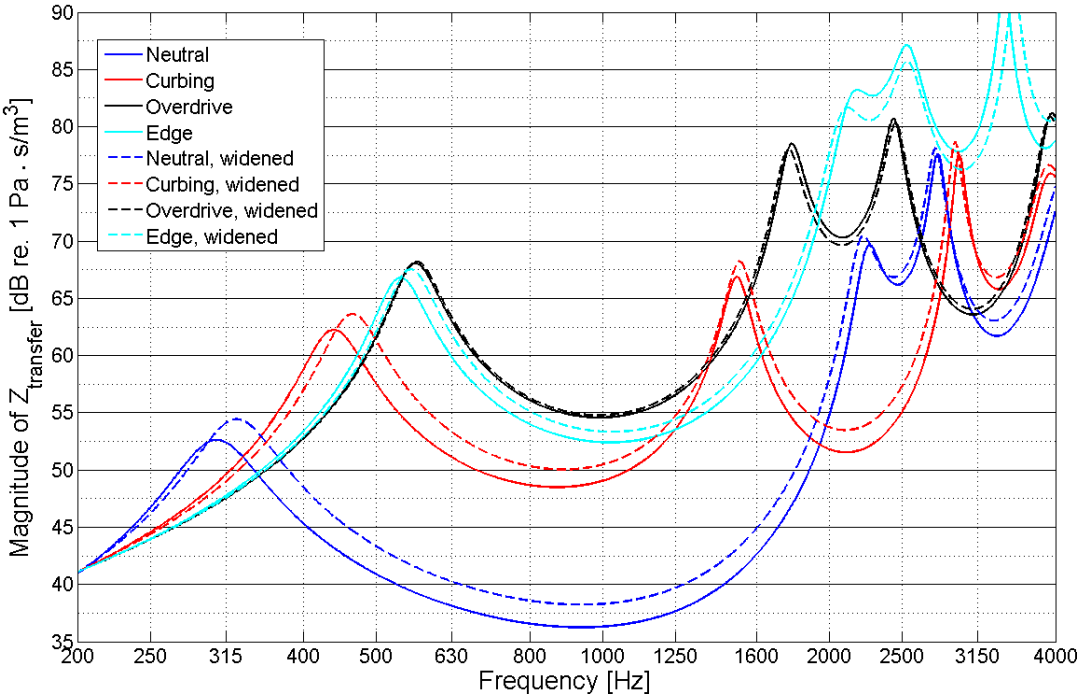
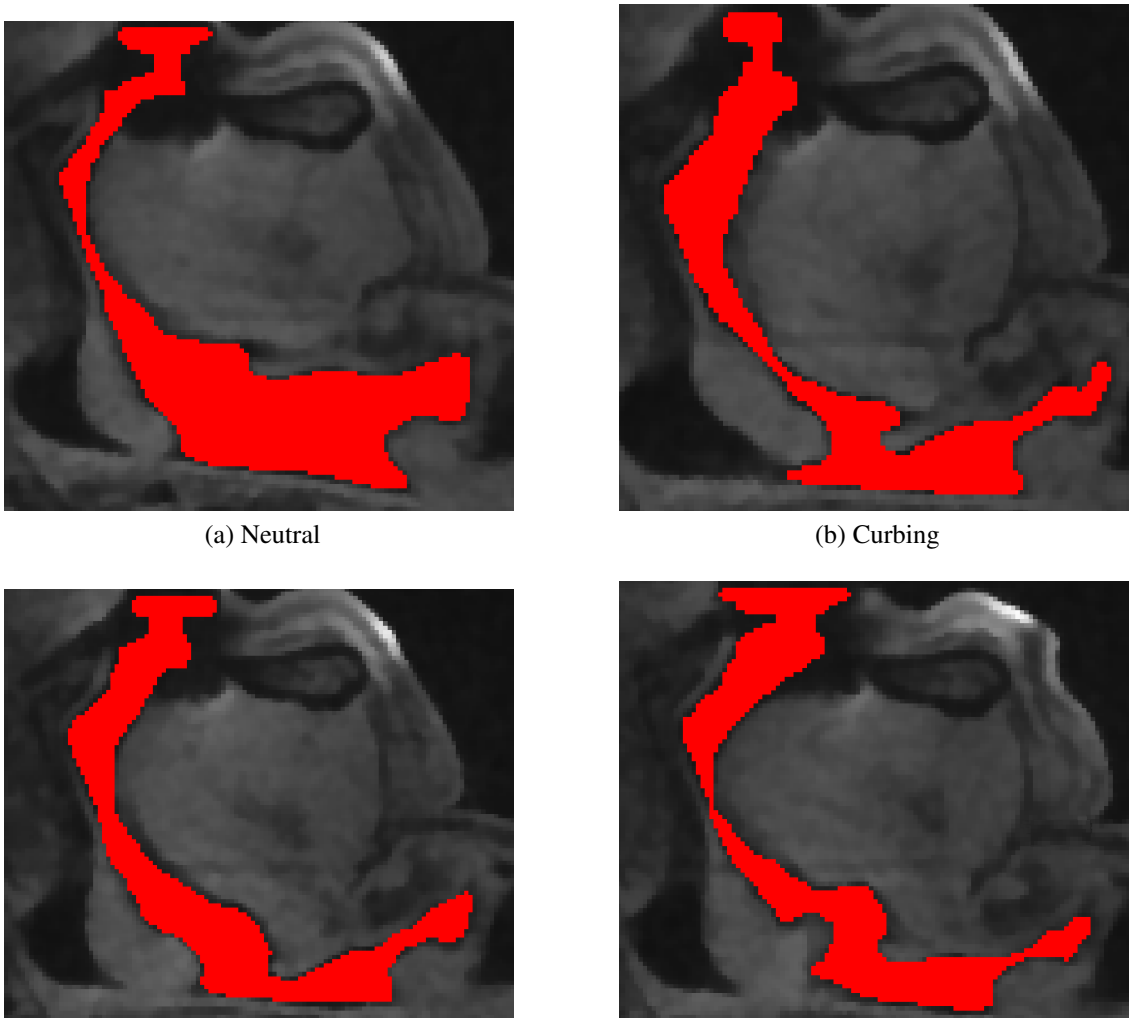


Figure 3.6: Calculated transfer impedance of the area-functions using the frequency-domain model.

The results show a rather large effect on the low frequencies of the modes that has a constriction at the back of the tongue. Even though this is a simulated experiment on the area function, the widened area function will be used in the analysis, because they fit the measured data better and there was no time to resegment the data.

The variations of the area-functions cannot only be attributed to the change of mode, but also to the change of vowels used. Overdrive and edge was principally phonated with the same vowel "EH", but neutral was with "EE" and curbing what with "OH". To be noted on the area functions is that the shape of the epilarynx tube is different for all the modes, the pharynx cavity varies with a constriction about 4 cm's from the glottis, which relates the the epiglottis. Overdrive does not have a constriction anywhere over the back of the tongue as all the other modes have, see figure 3.7. However, Neutral could be changed significantly when phonating another vowel. Curbing has a sharp constriction at the back of the mouth, where it is moved forwards in edge. All the modes have a horn-like opening of the area functions, where it is very prominent in Edge. This region is however a bit uncertain, since it is reconstructed.



(a) Neutral

(b) Curbing

(c) Overdrive

(d) Edge

Figure 3.7: The segmentation mask showing where the constriction appears

Another fact to emphasise is that in Curbing there is an opening to the nasal cavity. The area of the nasal cavity is not extracted in this study due insufficient MRI data. The opening may cause deviations in the results.

3.3 Numerical validation of the extraction

To validate the process extracting of the area-functions, the acoustics in the vocal tract is evaluated numerically using Comsol Multiphysics. In Comsol the propagation of acoustical waves is simulated in the cavity enclosed by the surface mesh with a radiation condition implemented at the mouth. All boundaries are rigid, simulating that vocal tract walls are hard. For this simulation the meshes without the piriform sinuses are used.

Setting up the verification

The procedure of the simulation is rather simple. The "Pressure Acoustics - Frequency Domain" analysis is chosen in Comsol, which is a type of analysis that is used to find steady state responses in acoustics of spaces and volumes [COMSOL, 2012]. The mesh is imported into Comsol where the radiation condition is implemented as seen in figure 3.8.

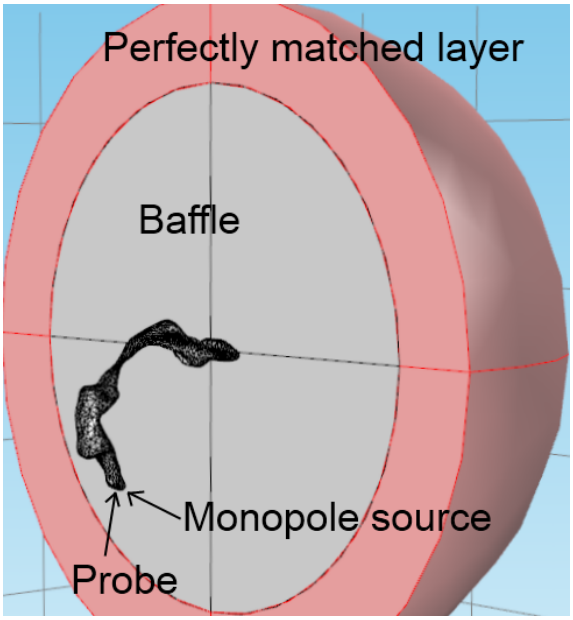


Figure 3.8: The setup in Comsol.

The baffled radiation condition is best modelled in Comsol by actually drawing up the baffle. The propagation is calculated in the inner halfsphere. The outer shell of the halfsphere is applied as a perfectly matched layer, which model anechoic conditions. Air is applied as the material of the volume and a monopole point source and a probe is placed at the vocal folds. Comsol then meshes the volume inside of the surface mesh into tetrahedron mesh elements (triangular pyramids) and solves the three dimensional wave propagation in all elements for a number of predetermined frequencies. These are set to the range from 100 Hz to 4 kHz with intervals of 10 Hz. The acoustical pressure recorded at the vocal folds maps out the impedance, as the monopole source delivers a steady volume velocity.

The numerically simulated input impedance is compared to the results from the frequency-domain model without losses, described in section 2.1. As mentioned in the previous section, the numerical solution was used to fine-tune the extraction of the area-function through the placement of the trail through the glottis. In the results of Overdrive the long and short trail is displayed for the illustration of the effects of the trail placement.

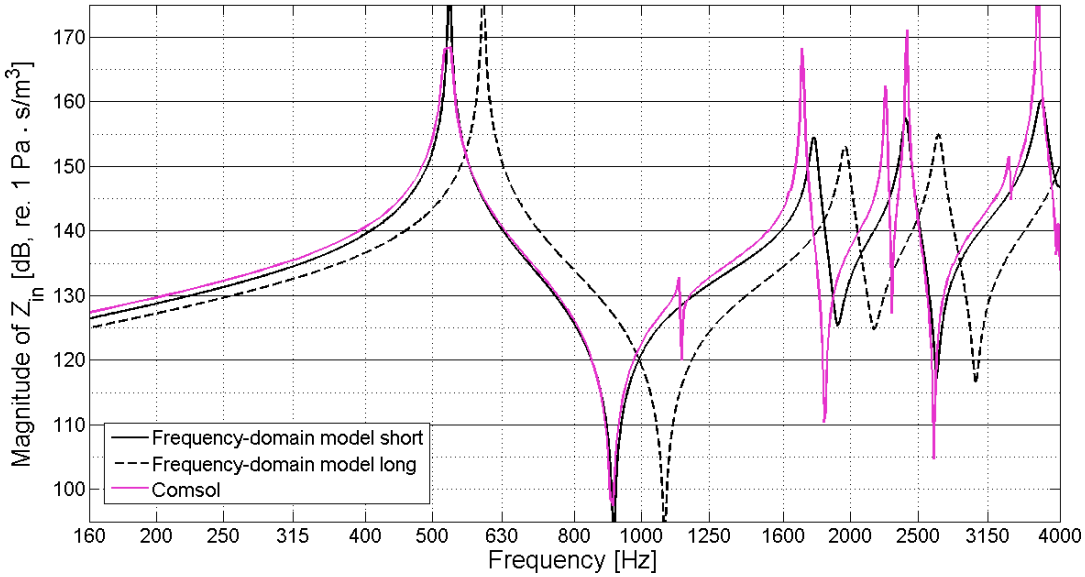


Figure 3.9: Verification of the extraction of the area-function for Overdrive: Comparison between results from the frequency-domain model and Comsol simulations.

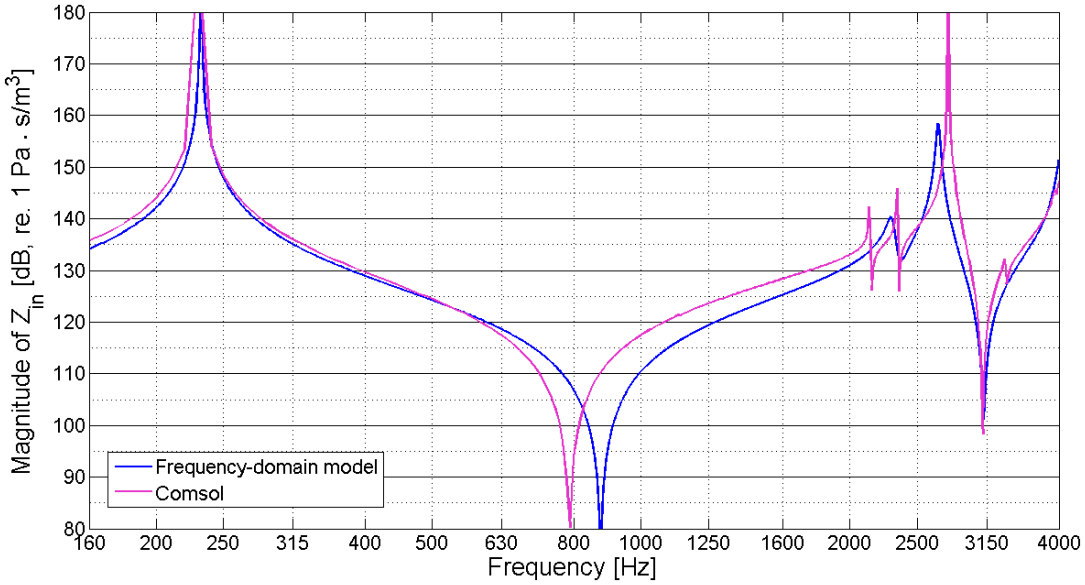


Figure 3.10: Verification of the extraction of the area-function for Neutral: Comparison between results from the frequency-domain model and Comsol simulations.

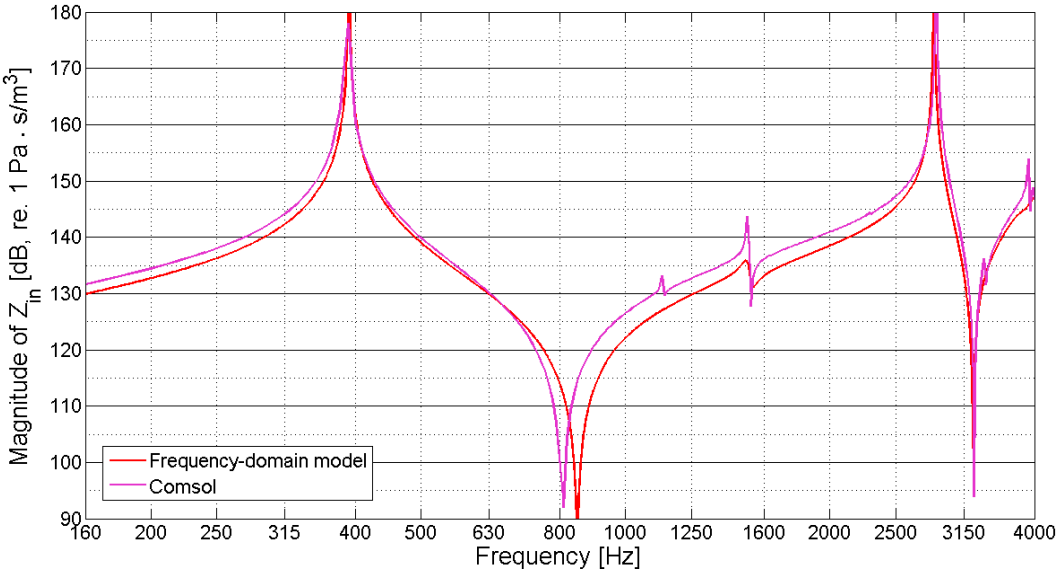


Figure 3.11: Verification of the extraction of the area-function for Curbing: Comparison between results from the frequency-domain model and Comsol simulations.

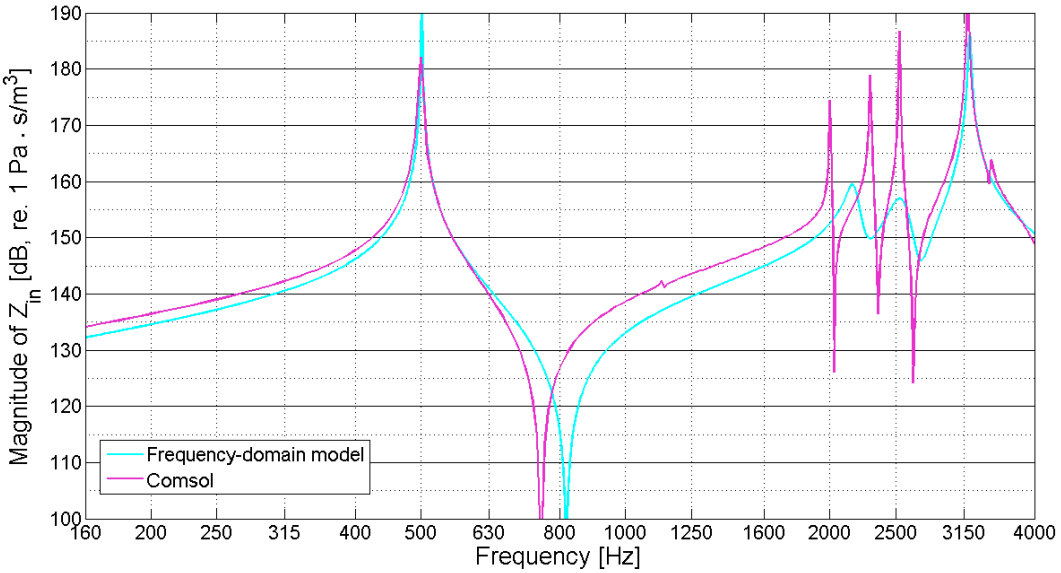


Figure 3.12: Verification of the extraction of the area-function for Edge: Comparison between results from the frequency-domain model and Comsol simulations.

The results generally show good agreement between the two models. The area-functions have been tuned with the purpose of matching up the whole impedance, but it was not possible to adjust for all characteristics. It is clear to see that the long trail is the more correct choice for the extraction of the area-function. The anti-resonance at around 800 Hz was only matched up in overdrive. Besides that, there is only slight deviations at higher frequencies, where some characteristics are missing in the area-functions. This is due to the simplifications of the uniform tube. The alteration to the area functions described in the previous section was done prior to this study.

CHAPTER 4

Experimental validations

To link the modelling of the area-functions directly back the modelled system, an experimental validation was performed using two measurement techniques. It is important to perform the validation to get the results as close as possible to the real values. A slight change in the estimated resonance frequency of the system could have the impact that a harmonic is above the resonance frequency instead of below it, putting it in compliant load instead of inertive load. The following experiments are performed:

- A direct measurement of the vocal tract impedance using the Vocal Tract Measured Impedance method (VTMI)
- Formant analysis performed on acoustic recordings.

The output of these two methods is mainly the placement of the formants on the frequency axis, but the impedance from the VTMI method can be compared directly with simulations.

The basics of the VTMI method is that the vocal tract is exited through the lips with a known volume velocity. The pressure response is recorded at the lips and from this an input impedance to the system seen from the mouth can be derived.

The formant analysis is performed using the software praat from [Phonetics Science, 2010]. It uses a Linear Predictive Coding (LPC) method to extract the formants from a recorded audio signal.

Both measurements were conducted the day after the final scanning session, which means that the elapsed time may cause that modes were sung differently.

In the final part of this chapter the results are compared to the simulated transfer functions, to finalize the validation of the area-functions.

4.1 VTMI

The method is described in Epps et al. [1997] and Kob and Neuschaefer-Rube [2002], but the measurement method used here is also inspired by a method to measure the greens function in a room [Jacbosen, 2011].

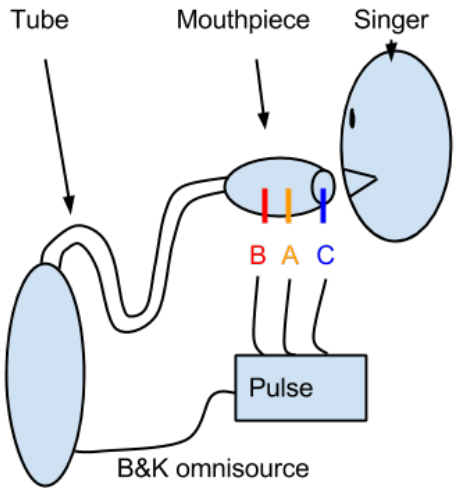


Figure 4.1: Illustration of the measurement setup of the VTMI method.

As seen from the measurement setup in figure 4.1, the B&K omnisource with a mouthpiece is used as a source. The mouthpiece is equipped with 2 pressure microphones, which makes it possible to measure the volume velocity through at finite difference approximation of the pressure gradient. This assumes that the sound field in the tube is one dimensional. The frequency limit of the measurement was limited to 3200 Hz.

The Greens function describes the pressure response at a point in space, to the volume acceleration of a point monopole. In this method the Greens function is rewritten to an expression depending on the frequency responses between the three microphones, using the finite difference expression for the volume velocity of the source. To find the transfer impedance, the Greens function is integrated and normalised by multiplying with $j\omega\rho$. See appendix D for details about the method.

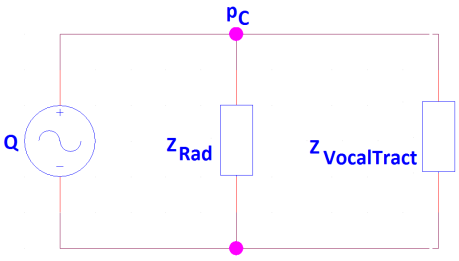


Figure 4.2: Schematic of the acoustical conditions at the mouthpiece

The acoustical situation at the mouthpiece is described by the acoustical circuit in figure 4.2. The volume velocity at the mouthpiece either enters the vocal tract or is radiated out

through the opening between the lips and the mouthpiece. The pressure response at the mouthpiece and lips p_c will therefore depend on the mouth input impedance to the vocal tract $Z_{\text{VocalTract}}$ and the radiation impedance Z_{Rad} in parallel.

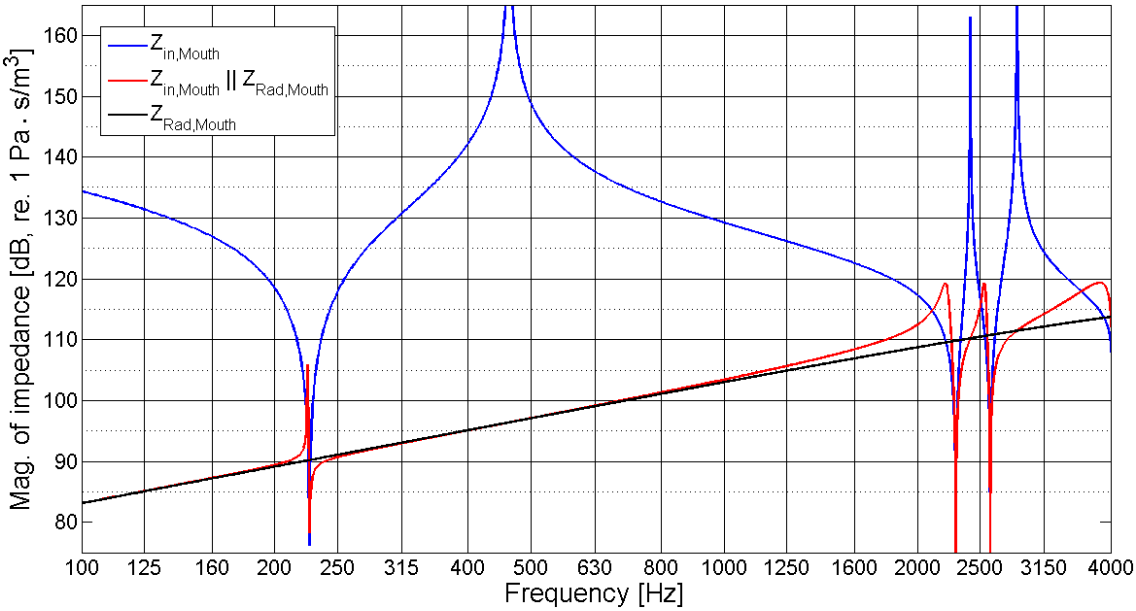


Figure 4.3: Simulation of the acoustical conditions at the mouthpiece in the VTMI measurement.

The results of a simulation of the total impedance is shown in figure 4.3, where Z_{Rad} is approximated by baffled radiation condition of the mouth opening and Z_{in}. This is simulated using the frequency-domain model without losses on the Neutral area-function. The simulation shows that the radiation impedance increases the frequency resonance peaks, while the anti-resonances are unchanged. This happen because the impedance of the vocal tract is much lower than the radiation impedance at the anti-resonance frequencies, which leads to most of the volume velocity flowing into the vocal tract. The actual vocal tract does not have as sharp anti-resonances and the dips will therefore not be as prominent as in this simulation, especially not at low frequencies where the radiation impedance is low.

A uniform closed tube has anti-resonance at odd multiples of 1/4 of the wavelength, which corresponds to the resonance frequencies of a uniform open tube. This concept is not immediately translatable to the vocal tract, but the results of the following simulation study in figure 4.4 shows that the anti-resonances of the impedance seen from the mouth, are comparable to the resonances of the impedance seen from the glottis.

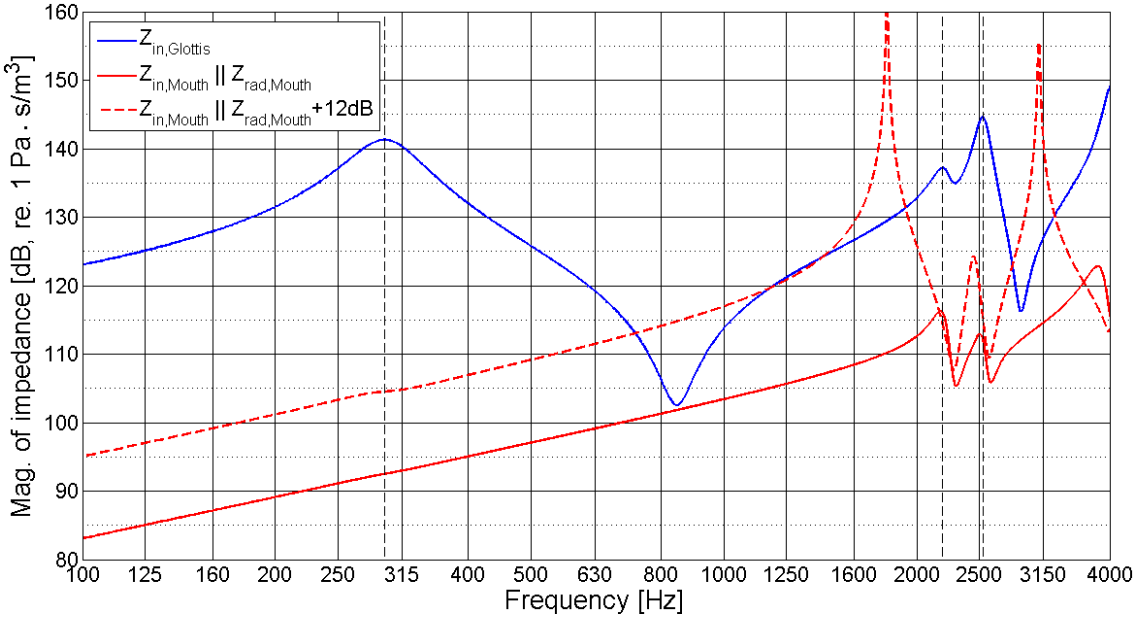


Figure 4.4: Results of a simulation study on formants and anti-resonances in the vocal tract. The black dotted lines are helping lines to compare the formants to the anti-resonances.

The simulation study is performed using the frequency-domain model with losses, using the extracted area-function of Neutral. The radiation condition was the baffled condition of the mouth, but a simulation with a 12 dB larger radiation impedance is presented as well.

The result shows that the anti-resonances does not exactly match up to the formants, see the black dotted lines. The anti-resonances overestimates the formants with 2-5% due to the radiation condition in the open vocal tract. A larger radiation impedance between the lips and the mouthpiece makes the anti-resonances stand out more, but the lowest antiresonance is only just noticeable on the curve. This complicates the prediction of the lowest formant, but the higher formant can be estimated directly within the given error margin. The measured impedances is also simulated for a direct comparison, but the value of the radiation impedance in the measurement is uncertain due to difficulties controlling the area of the gap between the mouthpiece and the lips.

The pulse system is used to control the measurement, which is set to measure the frequency responses between the three microphones microphones. The pulse system drives the omnisource with pseudo-random noise through an external amplifier. The phonation will create disturbances in the transfer functions, but because the method relies on the frequency response between the microphones, the disturbances are minimized.

The singer is asked to phonate in the four modes into the mouthpiece exactly as in the MRI scanner. The microphone is as close to the lips as possible, as the effect of radiation impedance is lowered by minimizing the gap between the singers lips and the mouthpiece.

Before the result of the measurement is presented it is important to note the circumstances in this experiment that may caused deviations:

- The singing technique may have been altered due to the obstruction of singing into a tube and trying to cover as much as the air gap around it. This created an abnormal placement of the lips and possibly a wider opening of the jaw. It was necessary to do this in the measurement to be able good results, as a result for F1 was aimed for. The experiment would have benefited from a mouthpiece that was fitted to the opening of the mouth.
- The measurement was done the day after the MRI recordings and the singer was standing with clear acoustical feedback of his voice. This may have caused the modes to change in phonation.
- In curbing the opening to the nasal tract can cause deviations, both in the measurements, but also from area extraction.

Results

The measured total impedance is compared with a simulation done on the mouth input impedance using the frequency-domain model. Some general characteristics match up between the measured and simulated mouth input impedance, but the radiation makes the results rather unstable when the input impedance is high. Therefore the results are better interpreted from a comparison of the formant positions, i.e. the anti-resonances in the data. The deviations are recorded from the plots and shown in table 4.1. The comparison is shown for Neutral in figure 4.5, but the remaining plots can be found in appendix D.

	F1 [Hz]	F2 [Hz]	F3 [Hz]
Neutral	430 (-16%)	2300 (-1%)	2650 (+5%)
Curbing	530 (-4%)	1280 (+28%)	2550 (+15%)
Overdrive	585 (+2%)	1550 (+18%)	2360 (+8%)
Edge	720 (-18%)	1900 (+14%)	2800 (-1%)

Table 4.1: Table displaying the derived formants. The deviations of the simulated result are shown in the parenthesis

An interesting note from the measured data, when the 2-5% overestimation from the radiation impedance is included, is that the first formant coincides with being:

- Just above the second harmonic in Curbing(522)
- Just above the second harmonic in Overdrive(522)

- Just below the third harmonic in Curbing(783)

Generally F3 had the best agreement with the modelled data. F1 is generally much lower in the modelling of all modes, except in overdrive. These data are used in the comparison with the modelled transfer function later in this chapter, it is important to note that a deviation of -2-5% is expected due to the methods.

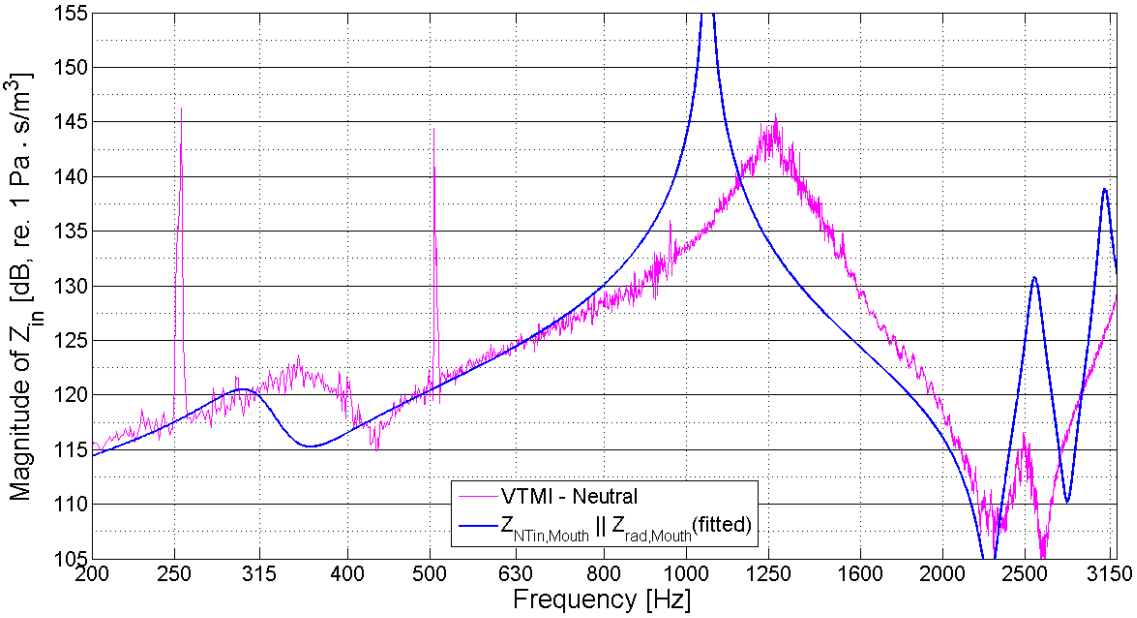


Figure 4.5: Result of the VTMI measurement for Neutral, compared to a simulation of the measurement.

4.2 Acoustic measurement

The Acoustical measurements are used for an analysis of the formants frequencies of the vocal tract, when phonating in the four modes. The recordings also provide information about the modes given the level of the harmonics in the signals, which is used later to compare to the modelling of the system.

The measurements were performed mimicking the conditions in the MRI scanner. Therefore the singer was phonating while lying on his back. In an attempt to disrupt the singers feedback a recording of the scanner noise was played back through headphones while he was wearing earplugs underneath. The measurements were performed in anechoic conditions, where the singers mouth was approximately 1 m away from the microphone. He was instructed to phonate the four modes in the same pitch and vowels that was used during the MRI recordings.

The signal was recording on a Sound Device 722 hard-disk recorder, through a DPA omnidirectional microphone. The sampling rate was 192 kHz and the resolution was 24 bit and the gain of the microphone pre-amplifier was the same through all measurements.

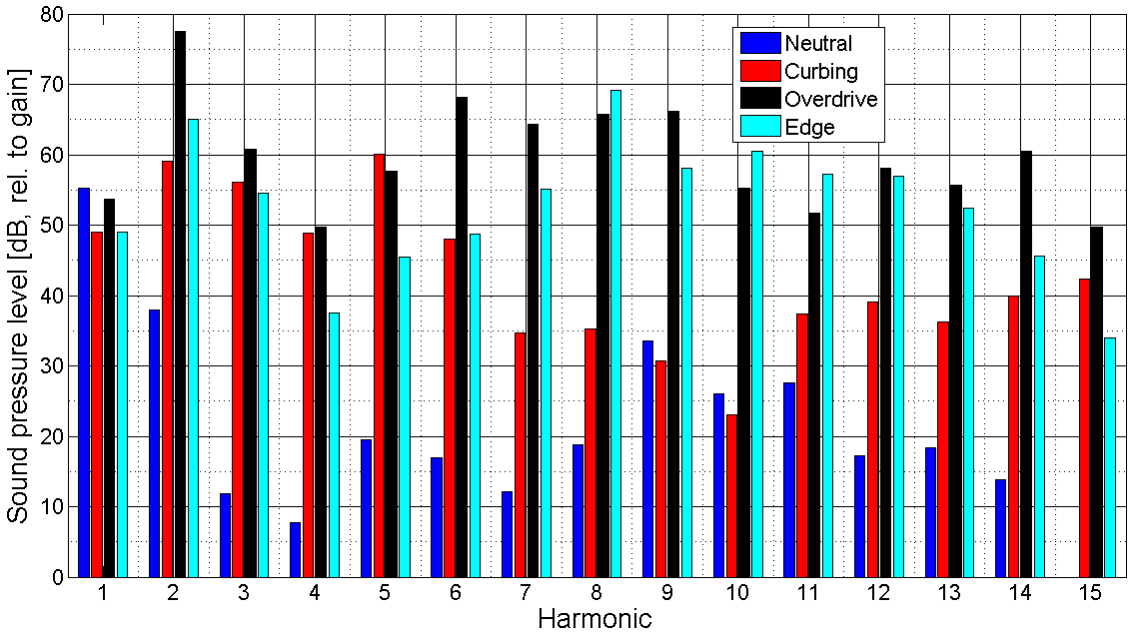


Figure 4.6: Strength of the first 15 harmonics of the four modes

All the recordings have been analysed and the data on the harmonics up to 4 kHz have been extracted, corresponding to up to the 15th harmonic. Figure 4.6 show the harmonic composition of the recordings. The deviation of the upper harmonics was quite large just throughout a recording of one phonation, so above 8th harmonic should be evaluated with precautions, the range up to 4 kHz is however used for the formant analysis. The

strength of these harmonics can be viewed as product of the glottal flow source spectrum and transfer impedance of the vocal tract. It is however clear that the metallic modes seem to be richer in overtones, but its not clear which mechanism that produces these.
In table 4.2 the A weighted sound pressure level of the harmonics are calculated [Jacobsen, 2010]. The A weighted values are not referenced to the absolute sound pressure since recorded pressure is relative to the gain of the pre-amplifier. It is a good measure for the loudness of the phonation, as the A weighting takes efficiency of the human hearing into account. From this result it is clear that the phonation is louder in the metallic modes.

From the harmonic composition of the signal it is possible to approximate the formant frequencies. This can be done by just analysing the harmonics for peaks, assuming the peak are there because the of a formant and have not originated from a strong harmonic from the glottal signal. This low-tech method has a maximum resolution of the spacing between the harmonic

Another method is to use a formant analysis algorithm that also finds the formants by analysing the harmonics. This analysis is performed here using a software called praat [Phonetics Science, 2010], which is a program used for analysis of phonation. It uses a Linear Predictive Coding (LPC) method that finds the poles in the filter and extracts the formant frequencies [Phonetics Science, 2010].The results are shown in figure 4.2.

	SPL [dB(A)]	F1 [Hz]	F2 [Hz]	F3 [Hz]
Neutral	40	324	2061	2707
Curbing	55	675	1306	2336
Overdrive	62	548	1685	2405
Edge	62	681	2053	2590

Table 4.2: Formants extracted using the LPC algorithm.

As the method was developed for determining the vowels in speech, the results should be take with precautions. The glottal flow waveforms of Curbing, Overdrive and Edge are especially expected to differ from the waveform in normal speech. An experiment was made to categorize how much a change in the glottal source will alter the predictions of the LPC method. A waveform was simulated using the transfer impedance of Edge and a harmonic dipole source with $F_0 = 261$ Hz and a volume velocity proportional to $1/f^2$ to approximate the glottal source [Fulop, 2011]. The source was made with and without a 6 dB boost of the 3rd harmonic. The LPC predicted the first formant correctly with the $1/f^2$ source, but overestimated it by 30 Hz for the boosted case. The higher formants were estimated correctly. In general F1 is not estimated very well by this method due to the spacing of the harmonics.

4.3 Comparison and validation

The prediction of the formants from the VTMI measurement and the LPC method have to be taken with precaution, due to the list of error sources of the VTMI measurement and the fact the LPC method was designed for analysing the speech signal. There are also many error sources through the scanning, image analysis, the extraction of the area-function and in the the simulation of the transfer function. An exact match may not be expected, but that necessarily does not mean that the prediction of the formant is wrong with regards to the modes of CVT.

The formants simulated in the transfer function used for the widening experiment, are used again to read off the formant frequencies. See table 4.3, where deviations relative to the measurements are calculated in pct.

	F1 [Hz]	F2 [Hz]	F3 [Hz]
Neutral, widened	**325(0%/-24%)**	**2220 (8%/-4%)**	**2770 (2%/5%)**
Curbing, widened	**465 (-31%/-12%)**	**1520 (16%/19%)**	**2940 (26%/15%)**
Overdrive, widened	**565 (3%/-3%)**	**1750 (4%/13%)**	**2450 (2%/4%)**
Edge, widened	**555 (-18%/-23%)**	**2110 (3%/11%)**	**2540 (-2%/-9%)**

Table 4.3: The results of the formant comparison. The pct. deviations are relative to: (LPC/VTMI).

The result shows that the predications of Neutral and Overdrive was fitted fairly good. Generally there are deviations between both the measurements, and the simulation, but it is rather hard to diagnose why a formant is deviating. It could be anything from a segmentation error or an error in the trail through the vocal tract, to an inaccurate measurement technique or a vowel that was sung a little differently. The underestimated first formant could be due to the constriction in the back of the mouth being segmented inadequately, but also a deviation could arise from the modelling of the yielding wall, since the pressure will be higher in this section due to the larger impedance of the small cross-section.

The simulation shows that the deviations between the measured and simulated formants are most extensive in the first formant and in curbing. The deviations in curbing may be the due to the unmodelled opening to the nasal cavity, given the results of modelling the nasal cavity in [Austin and Titze, 1997]. Another reason could be that the curbing modes simply was sung differently in the scanner, than in the measurements situation.

Summary and further analysis

Through this chapter the acoustics of the vocal tract have been investigated on the test-subject singing in the four modes of CVT in a pitch of C4 (261 Hz) and in four specific vowels. What lies outside of this pitch-area and vowel area cannot be accounted by this analysis. However there were some interesting points to conclude:

- The shape of the vocal tract has been characterised through MRI scans. The epilarynx tube was found to have a much smaller diameter in Curbing, Overdrive and especially edge.

  Neutral, Curbing and Edge have a distinct narrowing of the area at the back of the tongue. Overdrive did not have this narrowing.

- Through the VTMI experiment what appears to be a tuning of the first formant to the higher harmonics to harmonics was observed. This could also be a coincidence.

- The formants of the simulation of the transfer function matched up fairly well for neutral and overdrive, but the first formant of edge seemed to have been estimated too low. This could be due to an error in the segmentation, or perhaps the yielding wall, which should be modelled differently. Curbing was not estimated very well, however this could be attributed to the opening to the nasal cavity.

The observations that was done in the VTMI measurement about the tuning of the first formant is very interesting. However, the deviation between this observation and the results from the area-functions make it difficult to simulate these effects. However, it is still not clear whether or not the area functions are verified as corresponding to phonation within the given mode.

In spite of the deviations, all the modes will be simulated with the extracted area-function in the following section.

CHAPTER 5

Analysis and modelling

This chapter will begin with an analysis of the acoustic characteristics of the extracted vocal tracts, followed by the modelling of vocal fold vibration. The modelling will be performed on the extracted area-functions and the resulting characteristics will be compared to the harmonics from the recordings.

5.1 Investigating the effects of the vocal tract

In the literature the inertive load of the vocal tract is very often characterized as having a beneficial effect on the effectiveness of the glottal source. What happens under inertive conditions is:

- In the opening phase, the air flowing through the glottis is building up a pressure as it accelerates the mass of the air column in the vocal tract.
- When the glottis then starts the close and the flow decreases the air will keep its momentum moving forward, creating a drop in pressure above the glottis

This cycle is illustrated in figure 5.1.

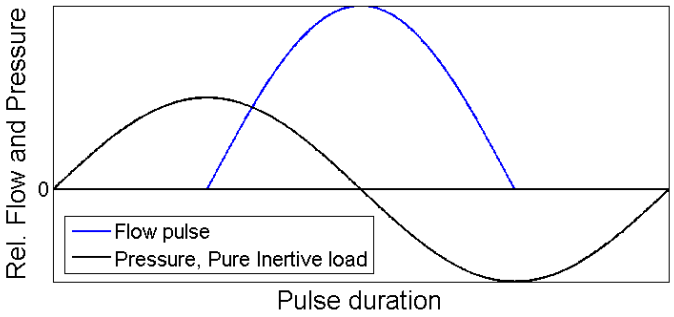


Figure 5.1: The cycle of pressure and flow during pure inertive load

When the pressure is coupled to the mechanical system, the pressure will push open the glottis in the opening phase and pull it shut in the closing phase. The flow through the glottis will be directly affected by the pressure drop across the glottis being smaller in the opening phase, and larger in the closing phase. The mechanism is the direct opposite when the load of the air column is compliant.

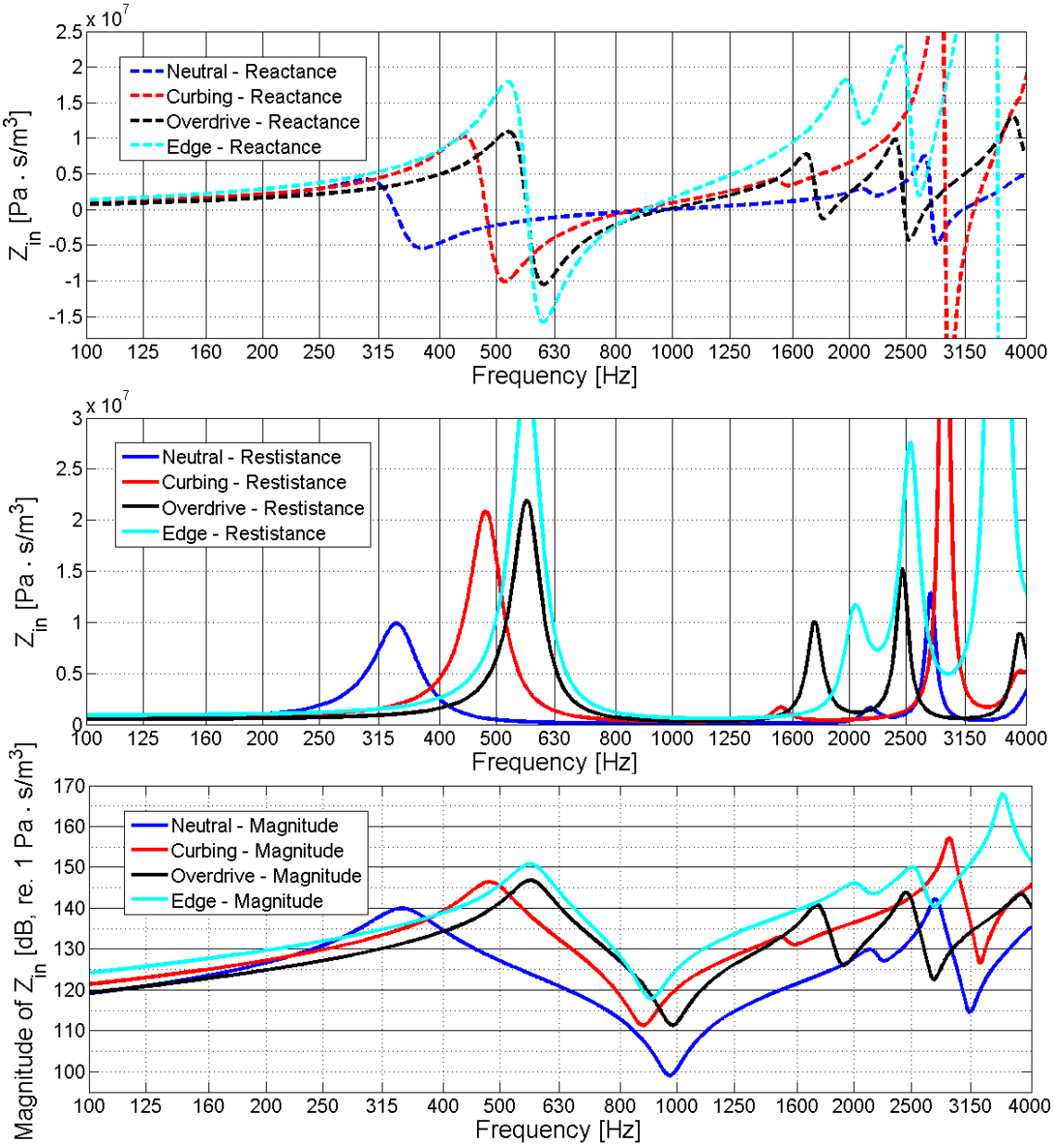


Figure 5.2: Input impedance of the vocal tracts

An inertive load is characterized by a positive reactance of the input impedance. In figure 5.2 the reactance, resistance and the magnitude of the input impedances to the vocal tracts are shown.

The resistance affects the phase of the impedance so that the pressure cycle shifts in the analogy described in figure 5.1. When the resistance takes over closer to the resonance, the pressure and the flow will be exactly in phase. This means that the pressure above the vocal folds always will be high when the glottis is open. This is not constructive for the source as the pressure will seek to push open the glottis during the whole cycle and the flow will be inhibited by the high pressure. At the anti-resonances frequencies the pressure will tend to be low, which implies the vocal folds will work as if no vocal tract was coupled to it. All the modes seem to be rather consistent about the antiresonance being located around 1 kHz.

According to Titze [2008], the load of a higher harmonic in a strongly coupled system can also affect the strength of that given source harmonic. Depending on the phase of this harmonic it can make the closing phase sharper, thus raising the energy of the harmonics [Backstrom et al., 2002].

Using the time-domain model, it is possible to monitor the pressure response to a flow pulse. As the flow pulse is not a sine wave that just includes energy at F_0, where the inertive region is described.

The vocal tracts have been extracted while the singer was phonating at 261 Hz. It is expected that the vocal tract is adjusted when the singer changes his/her phonation frequency, but for this phonation frequency we can see that:

- The first formant of Neutral is placed just above the phonation frequency and the second and third harmonic is within compliance region. At the first formant, the amplitude of both the reactance and resistance are not as high as in the other modes, but at the phonation frequency the reactance is actually 40% higher than in overdrive, see table 5.1.
- The First formant of curbing is a little lower than in Overdrive and Edge, which means the second harmonic has just shifted into the compliance region.
- Overdrive generally has the same characteristics as Edge, but has a significantly lower reactance and resistance at the phonation frequency than all the other modes.

These characteristics rely heavily on the vocal tract being correctly extracted. A wrong estimation of the first formant could change the load completely either at the phonation frequency or at one of the harmonics. The formant locations measured in the VTMI would alter the situation for Edge to accent the 3rd harmonic more, as F1 then would be just under this harmonic. However, most of the consideration about coupling is related to the fundamental frequency, as its response directly relate to the opening and closing phase of the pulse.

Mode	Neutral	Curbing	Overdrive	Edge
First harmonic [10^6Pa·s/m^3]	2.2 + i3.2	1.1 + i3.1	0.7 + i2.3	1.2 + i4.0
Second harmonic [10^6Pa·s/m^3]	0.1 - i0.2	0.9 - i1.0	1.0 + i1.0	1.6 + i1.8
Third harmonic [10^6Pa·s/m^3]	0.5 - i0.1	0.9 - i0.4	0.7 - i2.44	1.3 - i 2.7
Fourth harmonic [10^6Pa·s/m^3]	0.1 + i0.1	0.3 + i 1.28	0.3 + i0.5	0.6 + i 2.3

Table 5.1: Impedance data at the harmonics

The effects in the transfer functions

Using a harmonic source with a volume velocity proportional to $1/f^2$ Fulop [2011] in the transfer functions can tell us something about the accentuation that comes from the linear characteristics of the vocal tract. This approximates the source as hard, not exhibiting any of the coupling characteristics described above. This harmonic data will be used in the comparison to the model later in this chapter.

5.2 Modelling

The vocal fold vibration can now be simulated with the effects of the vocal tract configurations. The simulation study will use the vocal fold model and the time-domain model of the vocal tract described in chapter 2. Bear in mind that the simulation is a very rough approximation of the effects that happen in the system. Especially the accuracy of the vocal fold model can be improved. The goals for this simulation study is to:

- Explore the basic capabilities of the model.
- Gradually include the coupling to the vocal tract and explore the effects on the flow and the mechanics.
- Model the four modes using the extracted data for the vocal tract and the characteristics of the analysis in section 1.3
- Compare the results to measured data.

The input parameters to the model are:

- The area-function of the vocal tract.
- The resting position of the masses x_0 modelling the vocal fold adduction. For simplicity the value of this parameter is set the same for both masses.
- The subglottal pressure p_s
- The stiffness parameter q

The output of the vocal fold models is expected to be a train of flow pulses, but the following parameters will also be used to evaluate the performance:

- The flow u and the mean flow u_{mean}
- The minimum open area of the glottis A_min
- The open quotient O_q
- The movement of the masses x_1 and x_2
- The mean pressure created from the flow in the first section of the vocal tract (No reflections included)
- An intensity measure $|pu|$ derived from the mean flow and mean generated pressure (no reflections included)
- The pressure in the first section of the vocal tract.

Vocal fold oscillation without a vocal tract

The modelling part of this project will begin with the study of the basic function of the vocal folds model in order to establish a basis for interpreting the results. The model is capable of self-sustained oscillations without the load of the vocal tract. The model is first

used for a parametric study of the subglottal pressure p_s and the resting position of the masses x_0. See figure 5.3 and 5.4.

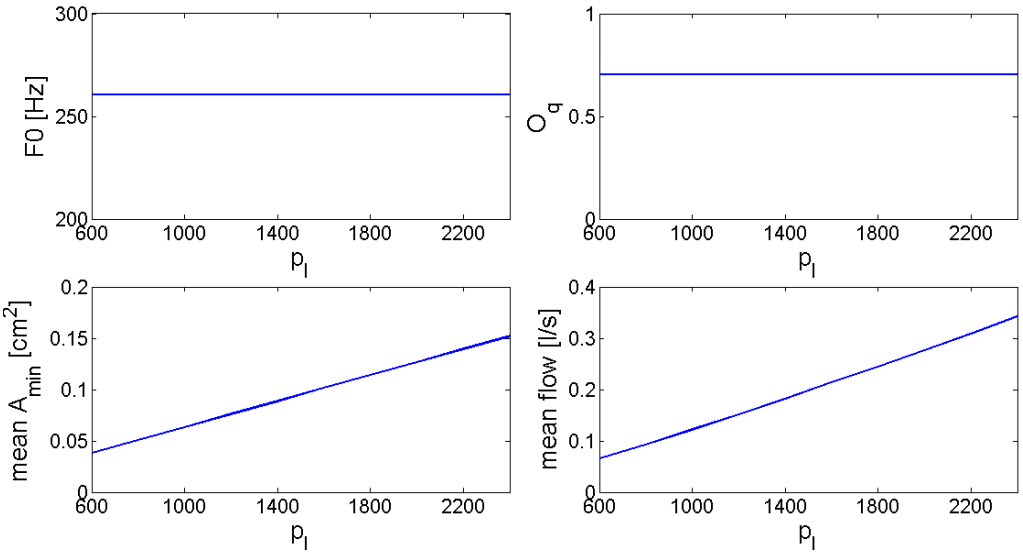


Figure 5.3: Results of the parametric study of the subglottal pressure (x_0=0)

The variation of the subglottal pressure shows that it only changes the mean area and mean flow linearly. The pressure may have other effects when more effects are introduced into the system, but the subglottal pressure will be kept constant at 1000 Pa in this study. This will most likely affect the output of the model, but it is done to limit the number of parameters to study.

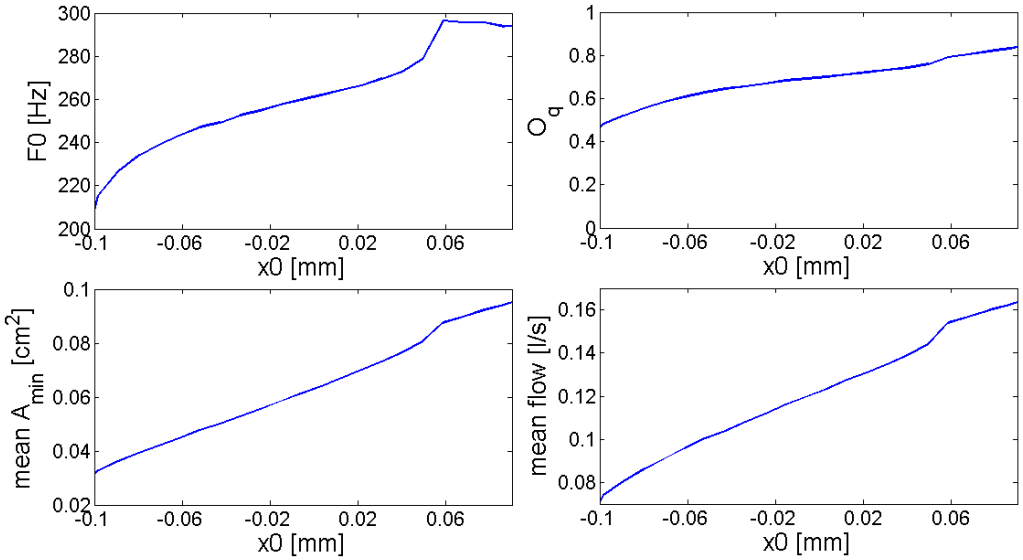


Figure 5.4: Results of the parametric study of the resting position of the masses (p_l=0)

The resting position of the masses show a more pronounced effect on the system, as this directly controls how the non-linearity of the collision is introduced. The phonation

frequency drops significantly when the resting position goes further into the negative region, as does the open quotient of the glottis. Phonation cannot happen below -0.1 mm at this particular tension parameter. The effect on the open area and mean flow seem to be of a more linear character.

It was hypothesised that Curbing, Overdrive and Edge is produced using pressed phonation, which is modelled in this study by letting the resting positions of the masses be negative. A pressed phonation and a normal phonation is modelled in figure 5.5. The drop in F_0 was adjusted back to 261 Hz by increasing the tension parameter. The tension when simulating pressed phonation was increased from $q = 2.0$ to $q = 2.6$, while the resting positions were changed from $x_0 = 0$ mm to $x_0 = -.092$ mm to obtain an $O_q = .49$.

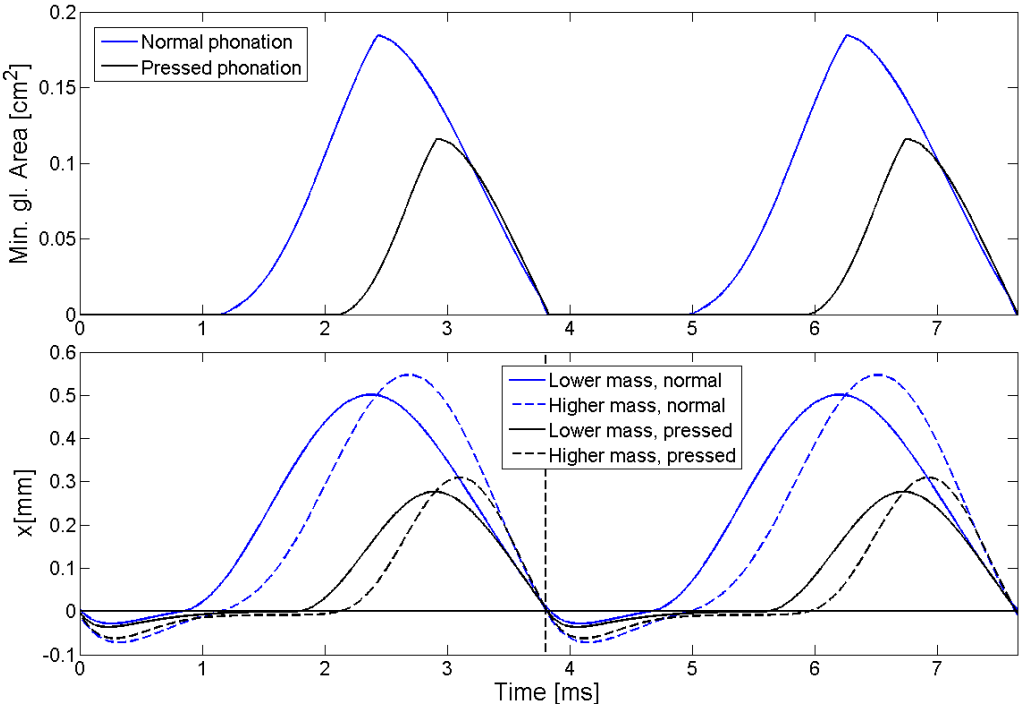


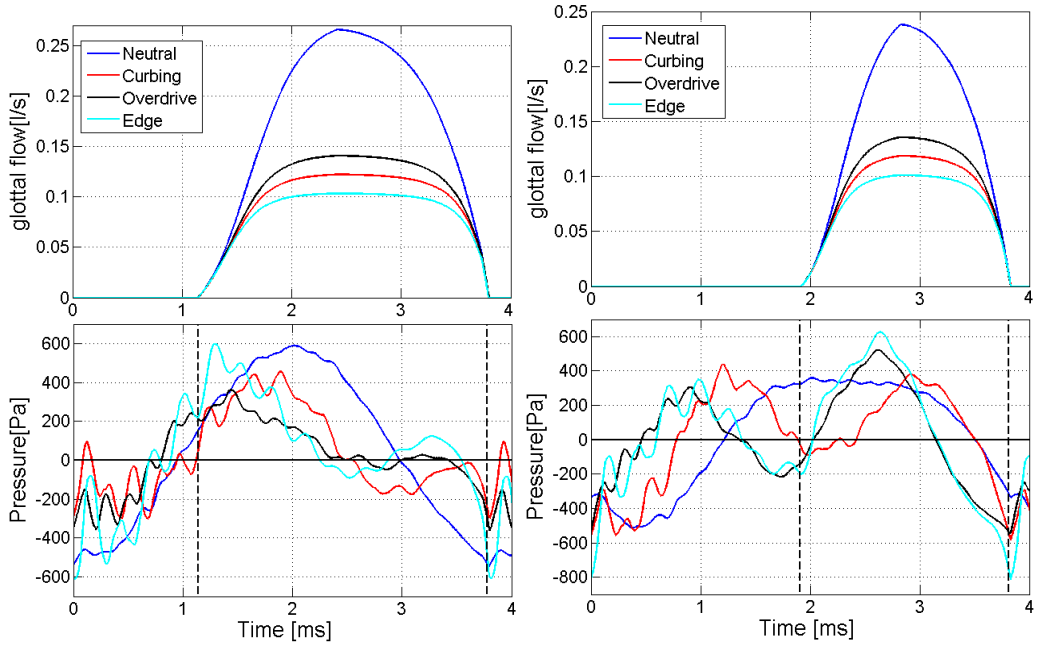
Figure 5.5: Normal and pressed phonation

Figure 5.5 shows that the minimum opening area pulse cuts off sharply in the top because of a change between which mass determines the minimum area. The plot of the mass movements show that the higher mass is lagging behind the movement of the lower mass. The lag between the masses is what causes the change between the diverging and converging glottis. The shape of the modelled waveforms of normal and pressed phonation look very similar. The shorter opening phase of the pressed phonation is the biggest difference, but it also has a slightly sharper closing phase, which relates to the strength of the harmonics.

Coupling to the vocal tract

The pressure in the vocal tract can be coupled back into the model of vocal folds in two steps - Flow coupling and Mechanical coupling. First the effects of no coupling is explored.

In figure 5.6 the flow pulse is modelled for the uncoupled case. The pressure in the figure is what is generated at the first section of the vocal tract. Table 5.2 shows the mean value data of the flow pulse.



(a) Normal phonation (b) Pressed phonation

Figure 5.6: The reflected pressure without coupling.

Mode	Neutral	Curbing	Overdrive	Edge
Mean flow, normal [l/s]	0.122	0.0665	0.0745	0.0577
Mean flow, pressed [l/s]	0.064	0.038	0.043	0.034
Mean pressure [Pa]	315	485	500	625
Mean pressure [Pa]	165	280	290	370
Intensity, normal [Pa l/s]	38.5	32.1	37.5	36.1
Intensity, pressed [Pa l/s]	10.6	10.7	12.2	12.4

Table 5.2: Data from the modelling of the vocal folds without coupling

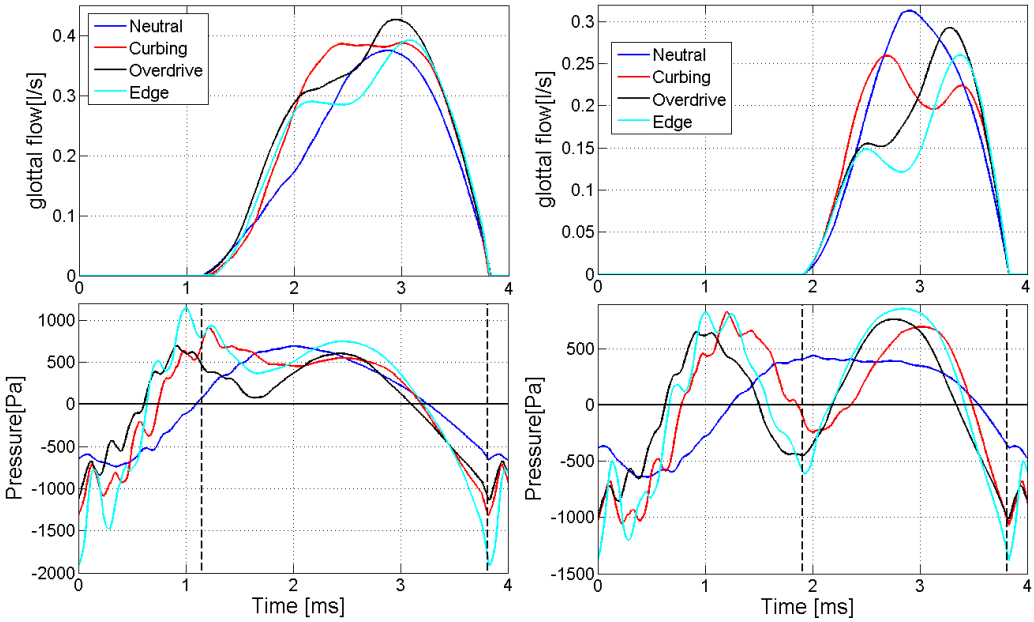
The vocal folds are under inertive load, as the positive pressure in the beginning of the cycle and negative in the end. This happens especially for the pressed phonation, where the timing of the pressure cycle is different due to the duration of the pulse. The fluctuations in the pressures are from the higher harmonics resonating in the vocal tract.

The pressure in Neutral is significantly different from the rest of the modes. It shows the inertive behaviour, but the pressure cycle seems to be shifted towards the closure. This is because the phonation frequency of Neutral is closer to the first formant, where the resistive load increases, see table 5.1. The pressure cycle in the other modes are under normal phonation rather alike, even though the pulse in Curbing also shifts a little towards the closure, but in pressed phonation the pressure for curbing shifted significantly. This is believed to be attributed to the compressive load of the second harmonic, see table 5.1, which shifts the phase of the pressure fluctuation. Overdrive and Edge are very alike, except that the pressure is a little larger in Edge.

Flow Coupling

The next step is to couple the reflected pressures of the vocal tract back into the flow model. This does not alter the movement of the vocal folds, but the effect of the coupling is that when the pressure difference across the glottis becomes greater, more air will flow through the glottis, given Bernoulli's equation.

A low pressure above the vocal folds will increase the flow, due to the greater pressure drop over the glottis. The feedback into the flow calculation will again increase the flow that was responsible for lower reflected pressure, increasing the flow even further. Over the first flow pulses the pressure in the vocal tract will build up and stabilise. The results of the flow and reflected pressure, while the flow coupling is active is shown in 5.7 and table 5.3.



(a) Normal phonation (b) Pressed phonation

Figure 5.7: The reflected pressure with flow coupling.

Mode	Neutral	Curbing	Overdrive	Edge
Mean flow [l/s]	0.147 (20%)	0.173 (162%)	0.177 (136%)	0.162 (184%)
Mean flow [l/s]	0.077 (20%)	0.067 (42%)	0.064 (49%)	0.054 (59%)
Mean pressure [Pa]	380 (20%)	1260 (160%)	1200 (140%)	1800 (188%)
Mean pressure [Pa]	200 (21%)	485 (73%)	435 (50%)	585 (58%)
Mean intensity [Pa l/s]	55.8 (45%)	218.4 (581%)	212.0 (465%)	283.9 (680%)
Mean intensity [Pa l/s]	15.3 (44%)	32.5 (200%)	28.1 (130%)	31.7 (155%)

Table 5.3: Data from the modelling with coupled flow, (%) show the change from the uncoupled case.

It is clear that the flow coupling has a significant effect on the flow pulse shape and amplitude. Generally the pulses have all been amplified, but the mean results of Curbing, Overdrive and Edge show that they have been amplified significantly. All the pulses have been skewed to the right, but the most significant changes happen in pressed phonation. The pulse of Overdrive and Edge is here very alike, but the top of the curbing pulse has been inversed. It still exhibits a strong negative pressure in the closing phase, but it is clear that the different load of the second harmonic is the factor here.

The normal phonation condition for Curbing, Overdrive and Edge gives an extreme increase of mean intensity, but a more moderate increase has been shown in pressed phonation.

Mechanical coupling

The final degree of coupling is the coupling to the mechanical system. Here the pressure above the vocal folds is taken into account in the calculation of the aerodynamic forces inside the glottis. In figure 5.10, 5.9 and table 5.4 the results of the flow pulses and reflected pressures are shown with both degrees of coupling activated.

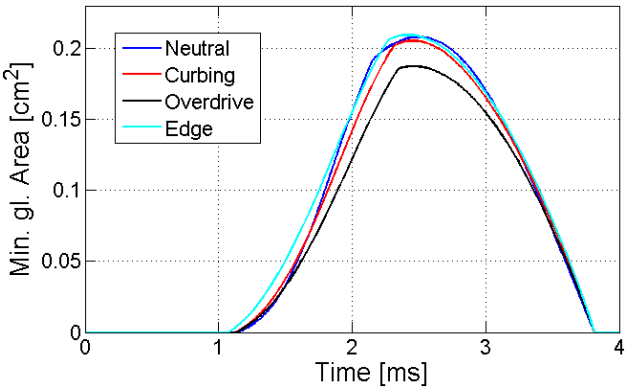
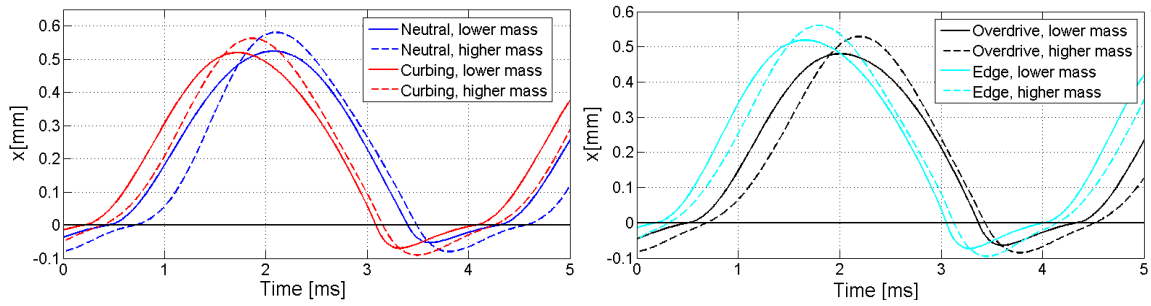


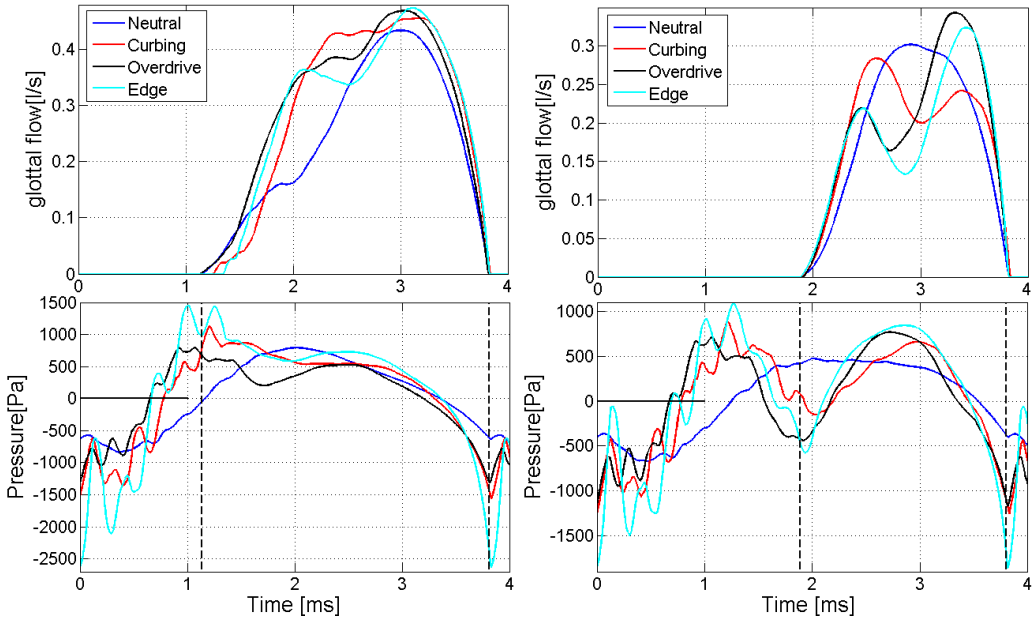
Figure 5.8: Minimum open area with mechanical coupling



(a) Normal phonation (b) Pressed phonation

Figure 5.9: The reflected pressure with flow and mechanical coupling.

Only the minimum area and mass movement of the pressed phonation is shown, since this is the most relevant for the strongly coupled modes. The effect of the final degree of coupling is very effected by the updated tension parameter, due to a change in O_q, see table 5.4. The vibration pattern of the masses is still the same. When comparing to the figure 5.5 it is evident that the relationship between the masses shift slightly earlier in all cases of mechanical coupling. The in the closing phase they seem to have the same relationship all the way down. and that the amplitudes have been amplified.



(a) Normal phonation (b) Pressed phonation

Figure 5.10: The reflected pressure with flow and mechanical coupling.

Mode	Neutral	Curbing	Overdrive	Edge
Tension parameter q	2.34(16%)	2.31(14%)	2.39(18%)	2.40(19%)
Tension parameter q	3.0(15%)	3.2(23%)	3.0(15%)	3.30(27%)
x_0 [mm]	-.087(-5%)	-.085(-8%)	-.085(-8%)	-0.083(-10%)
Mean flow [l/s]	0.167(13%)	0.202(17%)	0.207(17%)	0.201(24%)
Mean flow [l/s]	0.086(12%)	0.087(30%)	0.092(43%)	0.084(55%)
Mean pressure [Pa]	430(13%)	1470(17%)	1400(17%)	2185(21%)
Mean pressure [Pa]	220(10%)	630(30%)	625(44%)	900(54%)
Mean intensity [Pa l/s]	71.5(28%)	296.6(36%)	289.4(36%)	438.8 (55%)
Mean intensity [Pa l/s]	19.1(25%)	54.5(68%)	57.8(105%)	76.2 (140%)

Table 5.4: Data from the modelling with coupled flow and mechanics. (%) show the change from the uncoupled case.

The flow resulting from the addition of the coupling has altered the flow pulses, both Overdrive and Edge have a more distinct hump in the opening phase and Curbing has received one in the closing phase. Some of these changes may be related to the change in tension and adduction that was necessary to keep F_0 and O_q the same as without the coupling, but it makes sense that the tension of the vocal folds should be revised if the forces are different. The table shows that the mean flow, pressure and intensity again have increased compared to the results from the first degree of coupling, see 5.4.

5.3 Modelling the modes

The pressed phonation simulated in the previous section had an $O_q = .49$ for all vocal tracts. The phonation is now simulated with an O_q that matches the used vocal tract according to the data from [Sadolin and McGlashan, 2007]. See figure 5.11 and table 5.5 for the results.

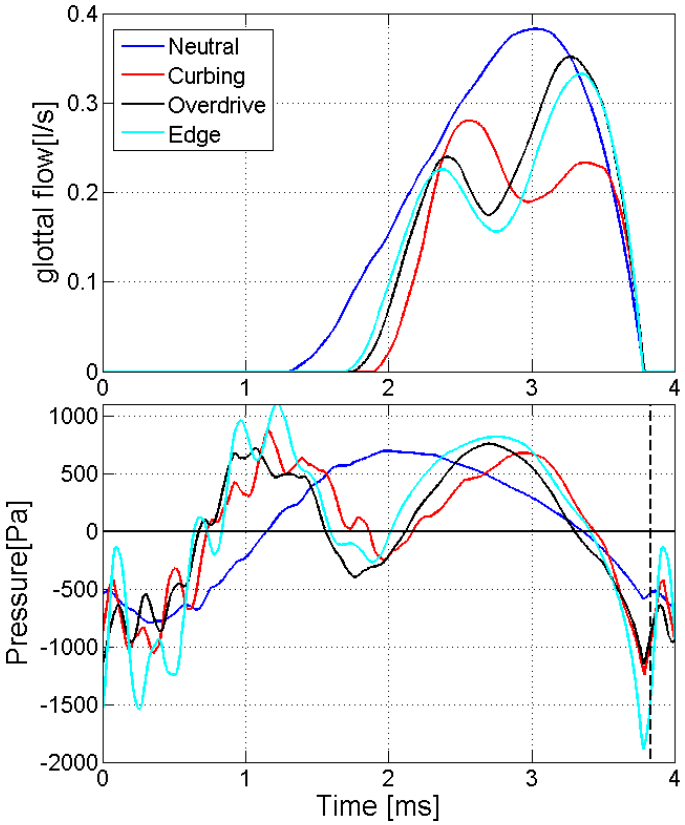


Figure 5.11: The modelled flow and reflected pressures.

Mode	Neutral	Curbing	Overdrive	Edge
O_q [l/s]	0.64	0.49	0.53	0.54
Tension par. [l/s]	2.4/2.09	3.1/2.39	2.75/2.26	2.95
x_0 [l/s]	-0.04	-0.09	-0.081	-0.08
Mean flow [l/s]	0.142	0.090	0.109	0.103
Mean pressure [Pa]	365	655	740	1115
Mean Intensity [Pa l/s]	51.9	58.9	80.6	114.7

Table 5.5: Data from the modelled modes with flow and mechanical coupling.

The larger open quotient for Overdrive and Edge does not seem to have changed the shape of the pulse, except for a longer duration.

To compare these glottal waveforms, the spectrum is calculated using DFT and the strength of the harmonics have been extracted. The source spectra are compared in figure 5.12. The specrum is weighted by f^2 for better comparison.

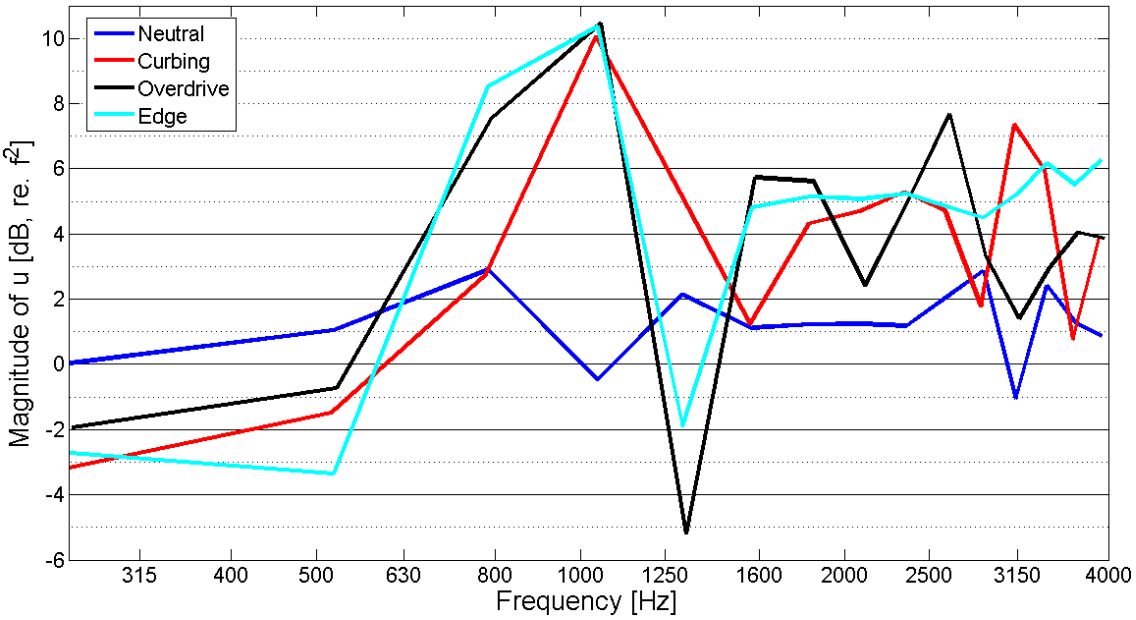


Figure 5.12: Level of the harmonics of the glottal flow, rel. to $1/f^2$

The spectrum of Neutral is very close to being relative to $1/f^2$. All the other modes exhibit a large increase around the fourth harmonic (1 kHz), which corresponds to where all the reactances have turned positive again. It was expected that the glottal source spectrum would have had much more energy at the 2nd harmonic for ovedrive and edge since this harmonic is in the inertive region below the resonance of the mode. The energy of the source at the fourth harmonic could be a product of having overestimated losses in the vocal tract model, which would be inhibit the resonances way to much.

Overdrive and Edge has many similarities, as could be expected from the flow pulse and the input impedance. They both have a large increase around the third harmonic. The spectrum at higher frequencies is generally around 4 dB higher than Neutral for all the other modes. This is believed to be originating from the steeper closing phase. The largest negative gradients have been calculated from the flow, characterizing the steepness of the closing phase, see table 5.6.

Mode	Neutral	Curbing	Overdrive	Edge
δu [m^3/s^2]	-1.13	-1.44	-1.63	-1.92

Table 5.6: Maximum negative gradients of the closing phase.

5.4 Discussion

The simulated harmonics of the glottal flow is used as an input for the transfer function of the vocal tracts using the frequency-domain model. This makes it possible to compare the results of the simulation to the measured data. This calculates the sound pressure levels of the harmonics as it would be in the measured setup. To scale the level of the simulated data and the measured data, F1 of Neutral has been scaled to the measured value. This is the reference for all the modes.

A comparison between the harmonics of the simulation, the measurement and a $1/f^2$ source with the transfer impedances is done for all the modes in the following figures. To accompany these figures are simulated sound examples (found in appendix). These have been synthesised using sine waves with the strength of the harmonics. Also the measured data has been processed this way for the comparison to be fair (original recordings are also included).

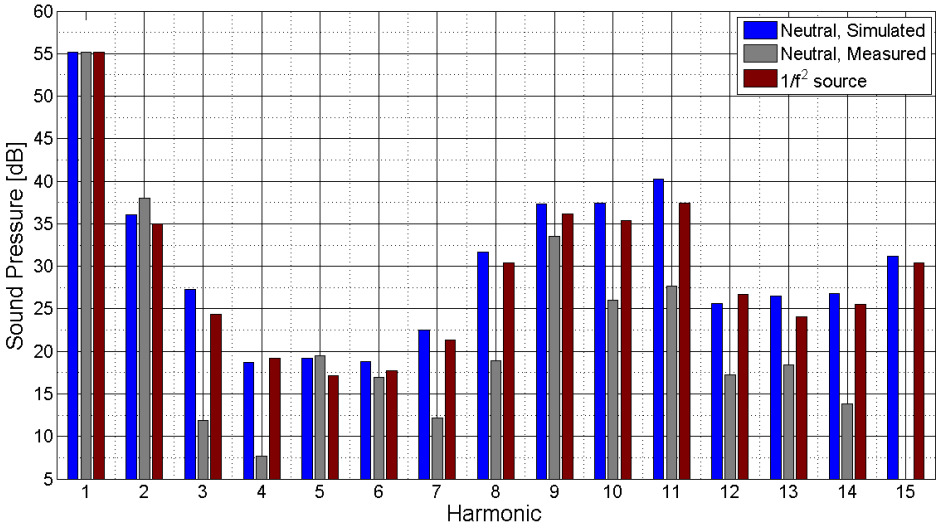


Figure 5.13: Comparison of the pressure harmonics, Neutral

It is clear that the overtones in the Neutral spectrum have been overestimated in the simulation and also the $1/f^2$ is overestimating. The problem may be that the model simulates the normal phonation with a full closure of the glottal flow. The third and fourth formant deviates both in the simulation and for the $1/f^2$ source. It may be due a too strong simulated coupling between the vocal tract and the source, since the vocal tract has an antiresonance at this position.

When listening to the recording of the real voice, then a slight breathy character is present, which may have been simulated better with a longer open quotient. It is very

clear from the simulated audio that the amount of overtones have been overestimated, as it sounds much "sharper". The chosen model cannot simulate the phonation without a full, rather sharp closure of the glottis, to simulate Neutral more accurately the model of the triangular glottis in [Birkholz et al., 2011] could be used. It would have been possible to characterize the open quotient better for the simulation by performing EGG measurements of the singer.

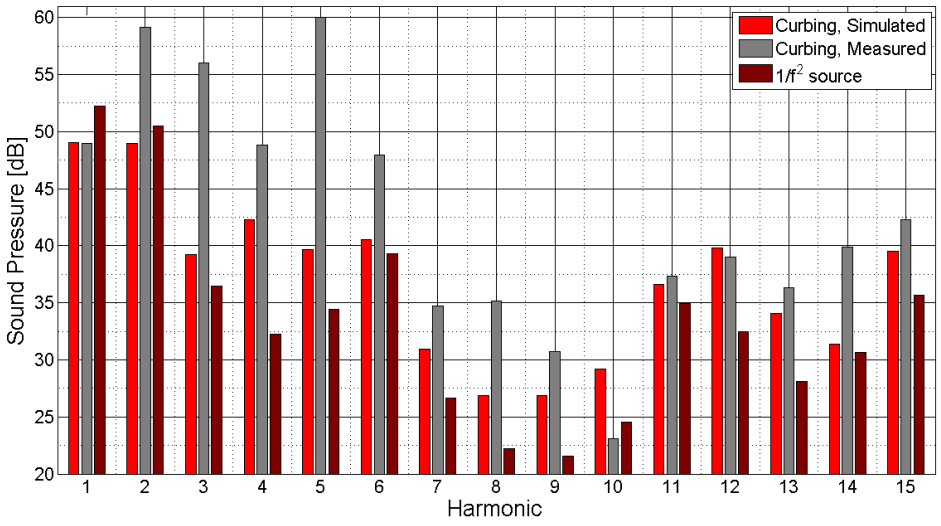


Figure 5.14: Comparison of the pressure harmonics, Curbing

Curbing has generally been underestimated and the third and fourth harmonic has is not estimated correctly given comparison with the measured data. It seems that a shift downward in F2, as was predicted by both measurements would align the harmonics more, but it does not change the fact that harmonic 2-6 has been underestimated.

When listening to the sound examples the vowels are not the same. However, the simulated Curbing also seems to be too sharp. There is a slight difference from the $1/f^2$ sound bite.

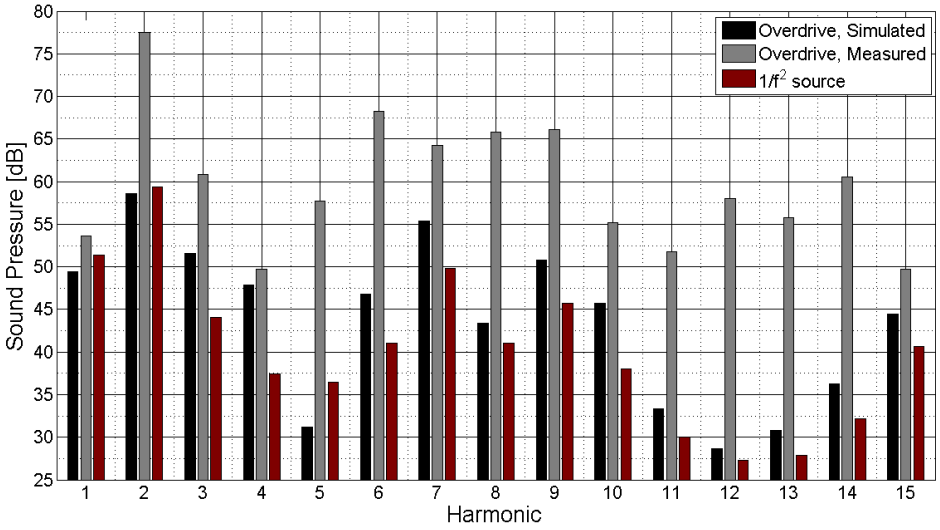


Figure 5.15: Comparison of the pressure harmonics, Overdrive

Overdrive has been generally very underestimated. The third and fourth harmonic seems to have been overestimated relative to the other harmonics. The second harmonic is underestimated very grossly, but in general 15-20 dB is missing in the simulation.

When listening to the sound examples it is clear that overdrive has not been modelled sufficiently. The vowel seems to be correct, but the timbre of overdrive is not there.

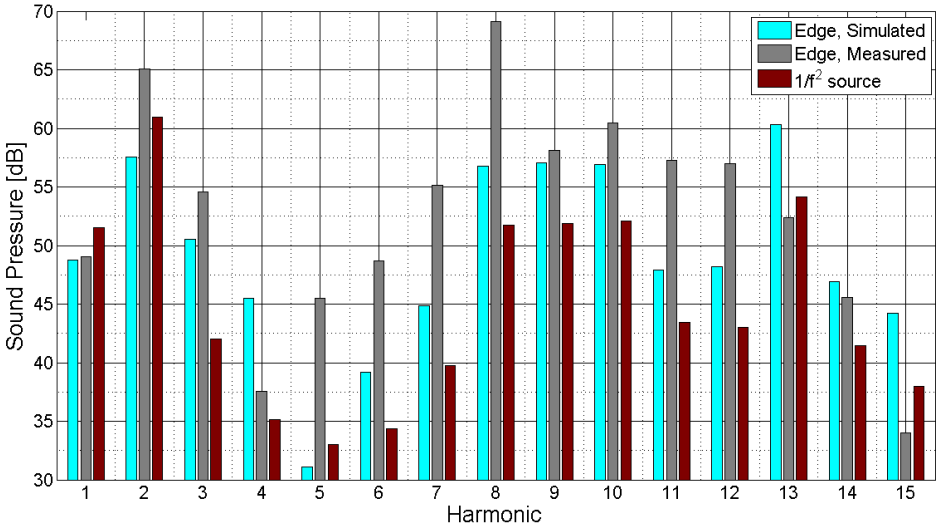


Figure 5.16: Comparison of the pressure harmonics, Edge

Edge is also underestimated but it did not seem to be as grossly as overdrive in the higher frequencies. There is also a misalignment over the first five harmonics

The sound examples reveals that the vowel and some of the distinct bite of edge is there, but it may just be because of the full closure that is modelled

Generally the modes did not compare very well with the measured data or how the expectations from the input impedance would be. The deviations is believed to be mainly attributed to:

- An exclusion of the subglottal pressure as a variable in the model. This would presumably have made the excitation of the vocal folds bigger which would lead to a larger flow, and a steeper closing phase. This was excluded from the simulation on the basis of the parametric study, where $x_0 = 0$. A negative resting position could have changed this parametric study.
- The losses of the vocal tract may have been overestimated in the time-domain model. This was a precaution that was taken for stability reasons which may have been to conservative.
- The exclusion of the subglottal duct may have been an oversimplification of the problem, in Titze [2008] it is described as contributor of inertive load to the system, but mostly with a larger supra-glottal area.
- The resting position of the masses was always set to the same value. An asymmetrical adduction could have been used to model another type of adduction where the setting of the vocal folds would have been more converging. Also a model for the muscle contraction is described in Titze [2002]. This could have been used to find a better relation between the masses than that in Birkholz et al. [2011].
- The full closure of the vocal folds did not give the correct roll-off for the harmonics of the source in the simulation. This is very evident in the simulated sound examples that all have a sharp sound to them. Once again the slightly more complicated model of the closure in [Birkholz et al., 2011] could perhaps be of use here.

The modelling was an attempt to simulate something complex with a very simple model and simplified input parameters. The result of this modelling is that it was not possible to simulate the characteristics of any of the metallic modes, but neutral could be approximated by what is characterized as normal phonation. The 2nd to 4th harmonic does however still show some deviation that is though to be attributed to the non-linear coupling. It was not possible to model the nonlinear source behaviour described in [Titze, 2008], where an inertive loaded harmonic can is amplified by the source.

By simply doing more iteration of the modelling, including more parameters than what was done in this project would perhaps make the model yield better results.

The between the extraction of the area functions and the measured data it was however rather evident what the source spectrum could look like. The second harmonic is very

strong in the measurement compared to the simulated vocal tract, which reveals that this should be strongly emphasized by the source.

The extraction of the area functions and the modelling of these was a partial success. Through the sound examples with the $1/f^2$ source it was very evident that correct vowel had been capture for Neutral, Overdrive and Edge. The vowel of Curbing was slightly off, but it is attributed the opening to the nasal cavity, which was not modelled. It is believed that also the first formant was estimated more or less correctly in Neutral and Overdrive.

The author hypothesised that a formant tuning was a benefactor of the characteristics of the modes. This was not confirmed through the modelling. A formant tuning strategy of the first formant of the modes in the lower part of the pitch range would make sense

CHAPTER 6

Summary and conclusion

The main focus of the project was to investigate the vibro-acoustics of the singing voice when using the singing techniques of Complete Vocal Technique - Specifically the different types of phonation within the technique, called modes. A similar project had been done prior to this at DTU, which was the starting point for this project [Selamtzis, 2011]. Even so, a rather extensive study of the physiology and research on the singing voice had to be done for the author to get familiarized with the subject.

The preliminary analysis of the modes pointed towards the coupling between the vocal folds and the vocal tract, being a major factor in the phonation and distinction between the modes. The possibility of performing MRI scans to measure the geometry was key for this project and a long time was spend on this process. The MRI scans were delayed by a consideration if the projects should be documented as a medical research project. That was however not the case.

An area functions for each mode was extracted from the acquired data, and the results were partly validated through three different types of measurement and simulation. One being a Comsol simulation on the extracted geometry. The other two were experimental methods that made it possible to approximate the formants of the vocal tract. the validation of the area functions for the modes Neutral and Overdrive was assessed to be succesful.

It was found that there was no constriction at the back of the tongue in overdrive, which is believed to have implications on the singers ability to tune the first formant higher up. This could explain the upper limit of overdrive, but it was not investigated.

The author believes that the VTMI measurement method can be used with more success than in this project if a mouthpiece is constructed to limit the radiation and not obstruct the singers technique. Through the VTMI measurements a hypothesis was set forth about formant tuning to achieve desirable non-linear effects from the source, but this was never verified.

The area function showed some interesting results of the size of the epilarynx tube.

The tube was about 0.5 cm^2 in cross-section for Curbing and Edge and around 0.3 for Edge. According to [Titze, 2008] this makes the interactions strong between the vocal folds and the vocal tract

After much research on modelling the vocal folds via lumped parameter models, the choice fell on a very simple two-mass model of the vocal folds and a wave reflection model of vocal tracts. Experiments with this model had been done prior to achieving the area function, but still some adjustments had to be made in the modelling phase.

Unfortunately there was not much time left of the project at this point and the first pass through the modelling did not yield successful results. The model was functioning, but the expected characteristics of the non-linearities did not show in the simulations. In the discussion a list of possible reasons for this has been listed. Through the sound example it was clear that some of that characteristics was in the vocal tract, something was missing, which must be the non-linear coupling.

This means that the goal of characterizing the modes of Complete Vocal Technique was not met

The MRI scans of the modes opens up to an overwhelming amount of possibilities of measurement, simulation and visualization of the singing voice. That is the main result of this project, there was just not enough time to do it all. Working with measurements on human body is a little different from working with the circuit boards and mathematics that is used to model them. A trained singer is a much more skilled user of his/her vocal instrument, than any model of it will ever be. A minute change in the balance of the dynamic system that makes a world of difference to the listener.

Future work

The work with the simple two-mass model has many possibilities for modelling the singing voice, however it is very important to have measurement to compare with these models, as the parameters tend to loose the relation to the physical vocal folds. The measurements could be:

- VTMI
- High speed video endoscopy, characterizing the vibrational pattern
- EGG measurements.
- Audio recordings

A study could be conducted on the formant frequencies of the CVT modes. It could be interesting to investigate the F1 tuning for the phonation of the modes.

The upper limit of overdrive is due to no constrictions being present at the back of the tongue, therefore removing an important tuning tool for the first formant.

Appendices

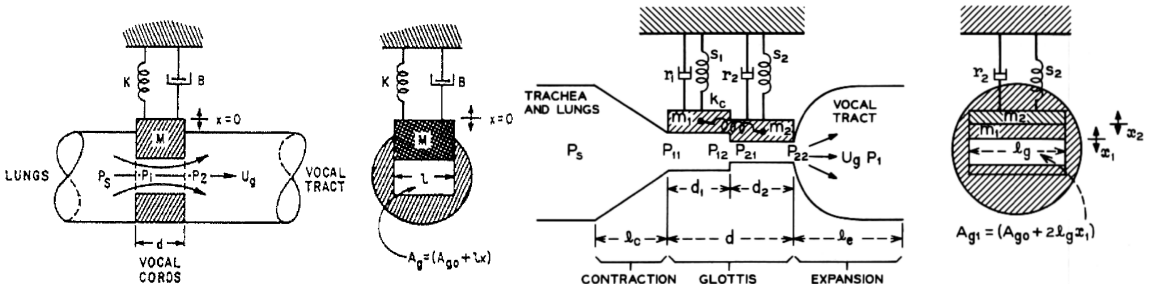
APPENDIX **A**

Review of vocal folds models

The following gives an overview over the advancement of lumped parameter vocal fold models.

Mechanical system

The simplest model of the vibration of the vocal folds consist of a single mechanical oscillator, modelling one vocal fold in [Flanagan and Landgraf, 1968]. The aerodynamic pressure and flow is modelled by Bernoullis eqation. This model has to be connected to a vocal tract to create self-sustained oscillation. See figure A.1a for an illustration of the model.



(a) Model of [Flanagan and Landgraf, 1968] (b) Model of [Ishizaka and Flanagan, 1972]

Figure A.1: Illustrations of the early vocal fold models

Implementing a second coupled mechanical oscillator can include the modelling of the shearing of the vocal fold cover in [Ishizaka and Flanagan, 1972], see figure A.1b. This model has the property that it can generate self-sustained oscillation through the change in the aerodynamics within the glottis during one cycle. This model includes non-linear springs and model the collision by an increased stiffness and damping in the collision region. A simplified version of this is used in [Steinecke and Herzel, 1995].

In [Story and Titze, 1995] the cover is modelled as a separate layer on top of body mass, thus introducing a third mass. The two cover masses are coupled to each other by a spring and connected to the body mass. This model is capable of modelling the shearing mode of the vocal folds more accurately.

More complex lumped parameter models have been used to describe the vibration of the vocal folds by adding more coupled masses.

An example of a two-dimensional lumped parameter model is found in [Adachi and Yu, 2005], that incorporate two dimensional movement of one mass.

In [Birkholz et al., 2011], a recently proposed model, the glottis is modelled as a triangular gap where the collision happens gradually when the vocal folds are abducted. This is done to capture the breathy quality at large abductions.

The rules for increasing the phonation frequency is usually a simple scaling of the resonance of the mechanical oscillators by the factor of Q:

$$\omega_0' = \omega_0 Q = \sqrt{\frac{kQ}{m/Q}} \tag{A.1}$$

In [Titze and Story, 2002], rules for controlling the parameters of the three mass body-cover model is implemented. The parameters are adjusted through three muscle activation parameters describing the degree of activation of the cricothyroid, thyroarytenoid and lateral cricoarytenoid.

The collision between the vocal folds is usually modelled by an increase of stiffness and damping in the mechanical system.

Aerodynamics and force distribution

The simplest model of the aerodynamics describe the change between the Bernoulli and the jet region abrubtly as the shape of the glottis changes between diverging and converging. See figure A.2 for an illustration.

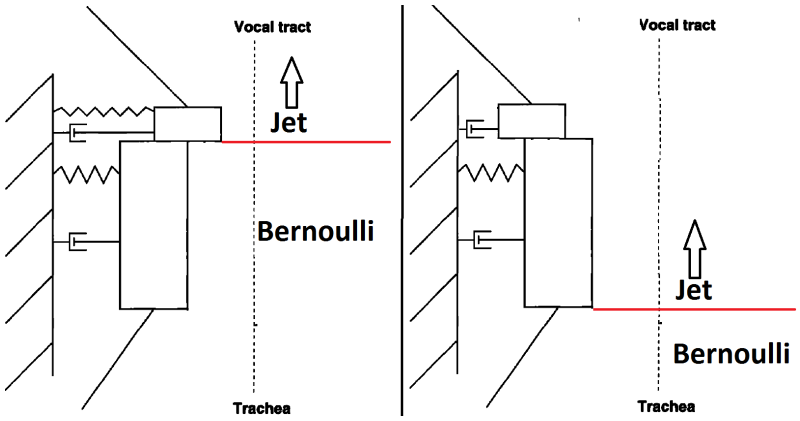


Figure A.2: Illustration of the aerodynamic situations in [Steinecke and Herzel, 1995]

The force on each mass is then calculated from the pressure situation across the glottis. The force only changes between the two regimes across the lowest mass, as the model only allows for Bernoulli flow in the region across the lowest mass. The force from the epilarynx pressure p_e is applied when in the jet regime. The distribution of the thicknesses of the masses are usually as in figure A.2 when this model of the aerodynamics is used. This model is used in [Steinecke and Herzel, 1995] and [Story and Titze, 1995].

An addition to the pressure distribution is made in [Birkholz et al., 2011], where the inlet and outlet of the glottis is included is included.

[Pelorson et al., 1995] and [Lous et al., 1998] use a more accurate model of aerodynamics. A jet separation point is defined as the glottis shifts between the converging and diverging shape. The jet separation point will move and define the size of regions where Bernoulli and jet flow are assumed. The force distribution is calculated by integrating the aerodynamic forces along the glottis thickness. The inlet and outlet of the glottis is also included in the force distribution. The thickness of the masses is more evenly divided when using this model.

A closed form calculation of the flow, done in [Titze, 1984] is very widely used in the more recent models. It uses the minimum glottal area and the lung pressure to calculate the flow, but also incorporates the coupling to pressure in the trachea and the vocal tract.

The [Ishizaka and Flanagan, 1972] model used a method of calculating the flow where both viscous and inertive terms of the flow are taken into account.

The pressure in the vocal tract affects the aerodynamics by altering the recovery pressure of the jet. This is incorporated in the closed form flow equation, but it also has an effect of the aerodynamic forces that affect the movement of the mechanical system.

APPENDIX **B**

Image analysis and creating 3d meshes

The flow of actions in creating the mesh can be found in the following figure (B.1).

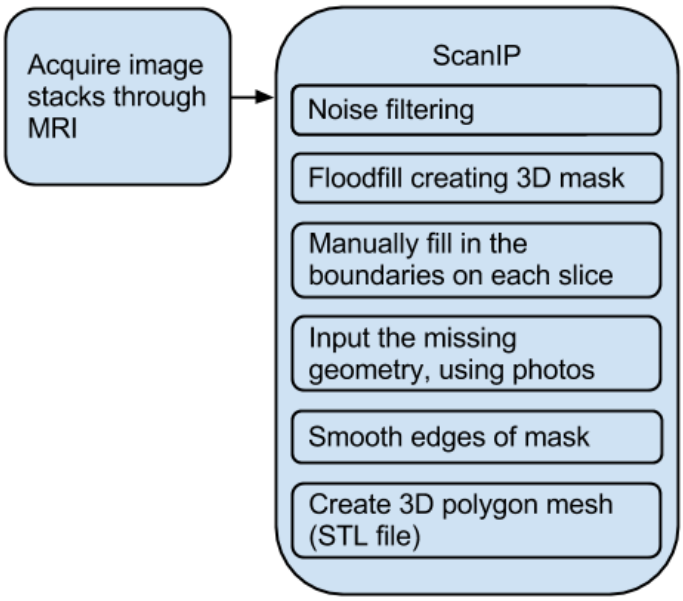
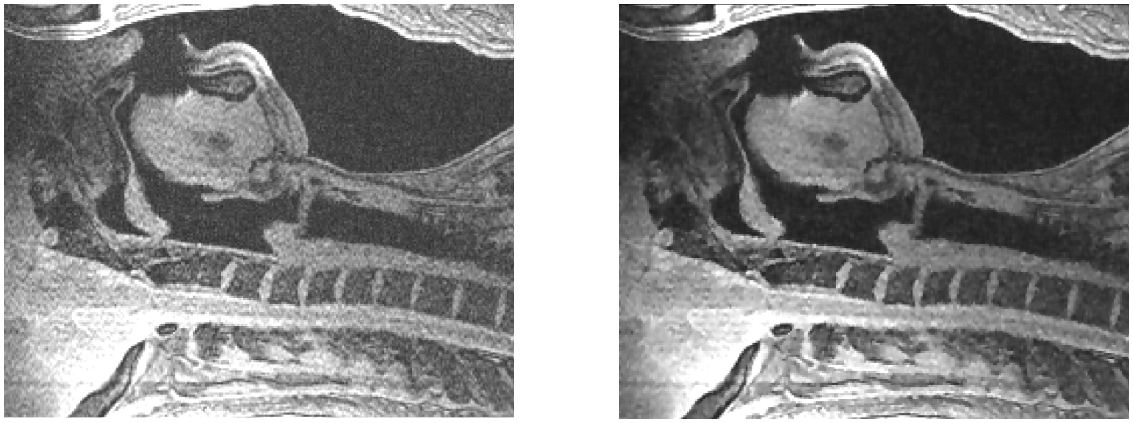


Figure B.1: Flowchart of the image analysis process

The recorded images was enhanced by using one of ScanIP's build-in noise removal filters. The filter that was used is a gradient anisotropic diffusion filter, which is an algorithm that removes noise while still preserving the edges in the image [Farsight, 2009]. The same filter was applied to all images with the same settings. In figure B.2 the effects of the filter are shown in a before and after illustration.



(a) Before (b) After

Figure B.2: Effect of noise reduction filter, illustration by before and after image

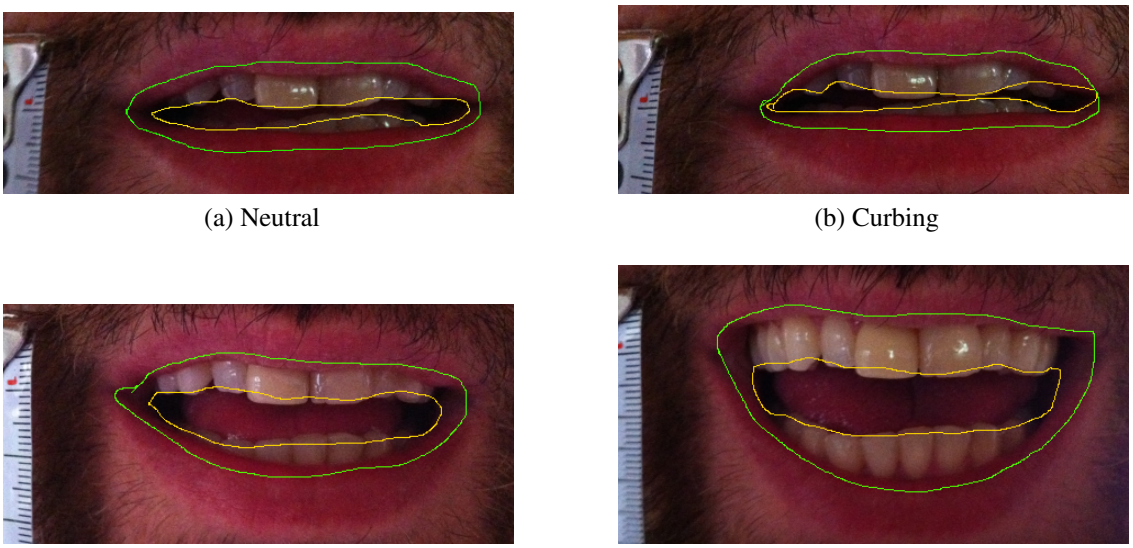
The segmentation of the airway was done using a floodfill command that creates a three dimensional mask over an interconnected region across the stack of images. The adjustment of a threshold command determines where on the boundary the segmentation will be, but it does not seem to be consistent consistent enough and the mask tends to flood across the boundaries. Therefore the floodfill was used at a low threshold and a manual outline of the airway was performed, ensuring a consistent segmentation.

Figure B.3: The mask of a slice in the saggital plane (Neutral mode), showing the segmentation of the airway

In the later analysis of the area it is found that image stack no. 3 (saggital plane) yielded comparable results across the whole vocal tract in all modes. Therefore these recordings

are used to produce a mesh that covers the whole vocal tract. A complete model is needed when doing the three dimensional numerical modelling and visualization.

Also on this series the lips, teeth and the tip of the tongue are reconstructed using photos shown in the following figure (figure B.4)



(a) Neutral

(b) Curbing

(c) Overdrive

(d) Edge

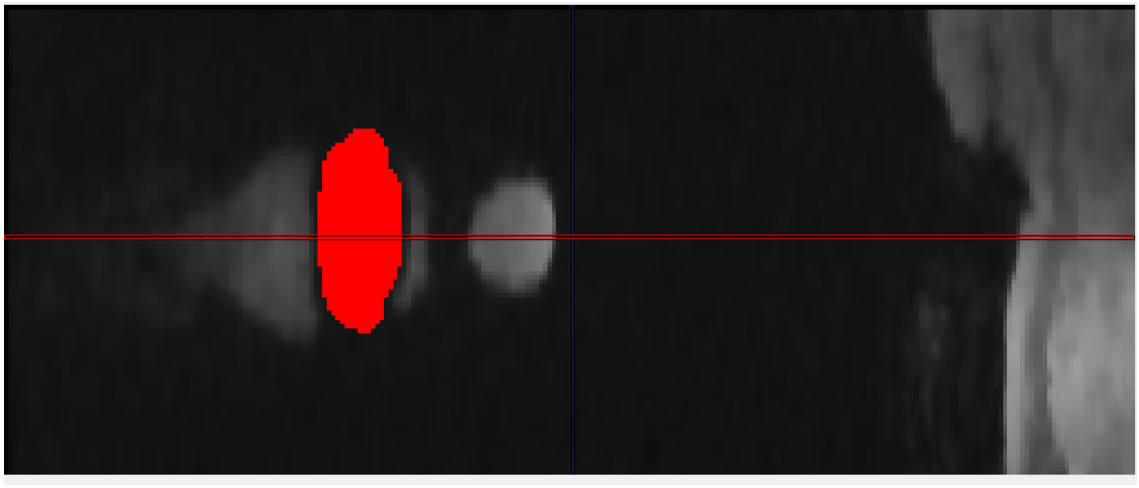
Figure B.4: Photos showing the opening at the lips and teeth in the four modes

A circumference of the lips and of the teeth is drawn up in MS paint, and then analysed in MATLAB to extract the area using the scale seen in the left of the picture, see table B.1.

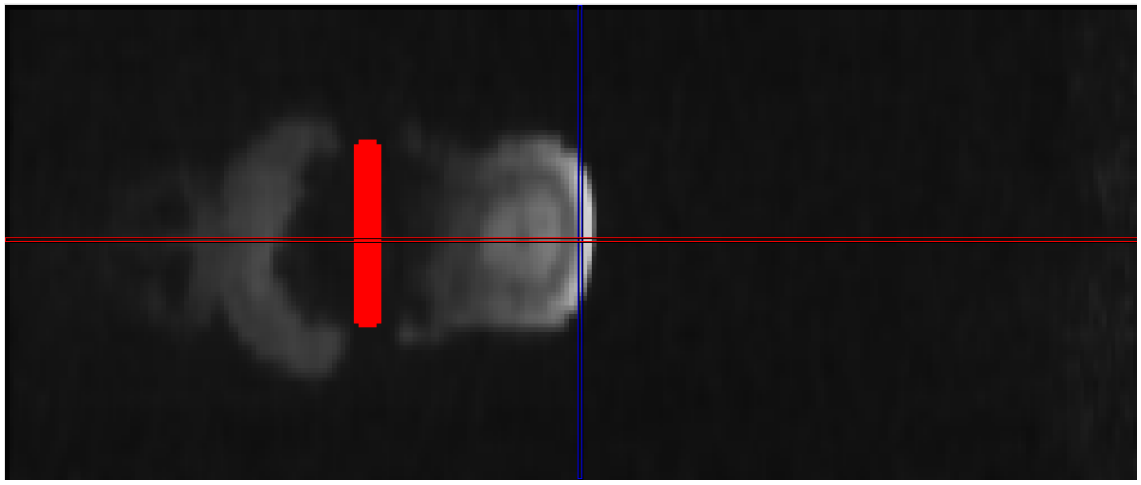
	Neutral	Curbing	Overdrive	Edge
Lip area [cm^2]	5	4.5	6	11
Teeth area [cm^2]	1.5	1	2	3

Table B.1: Lip and teeth area

The shape is reconstructed as best possible in ScanIP on the slice stack no. 4. The shape of the teeth and lips is simplified by being straightened out to a tube, i.e. simplifying the curved shape of the lips. The lips are somewhat visible in the recorded images, so both the shape from the photo and the recorded image is used for reconstruction. Screenshots can be found in figure B.5.



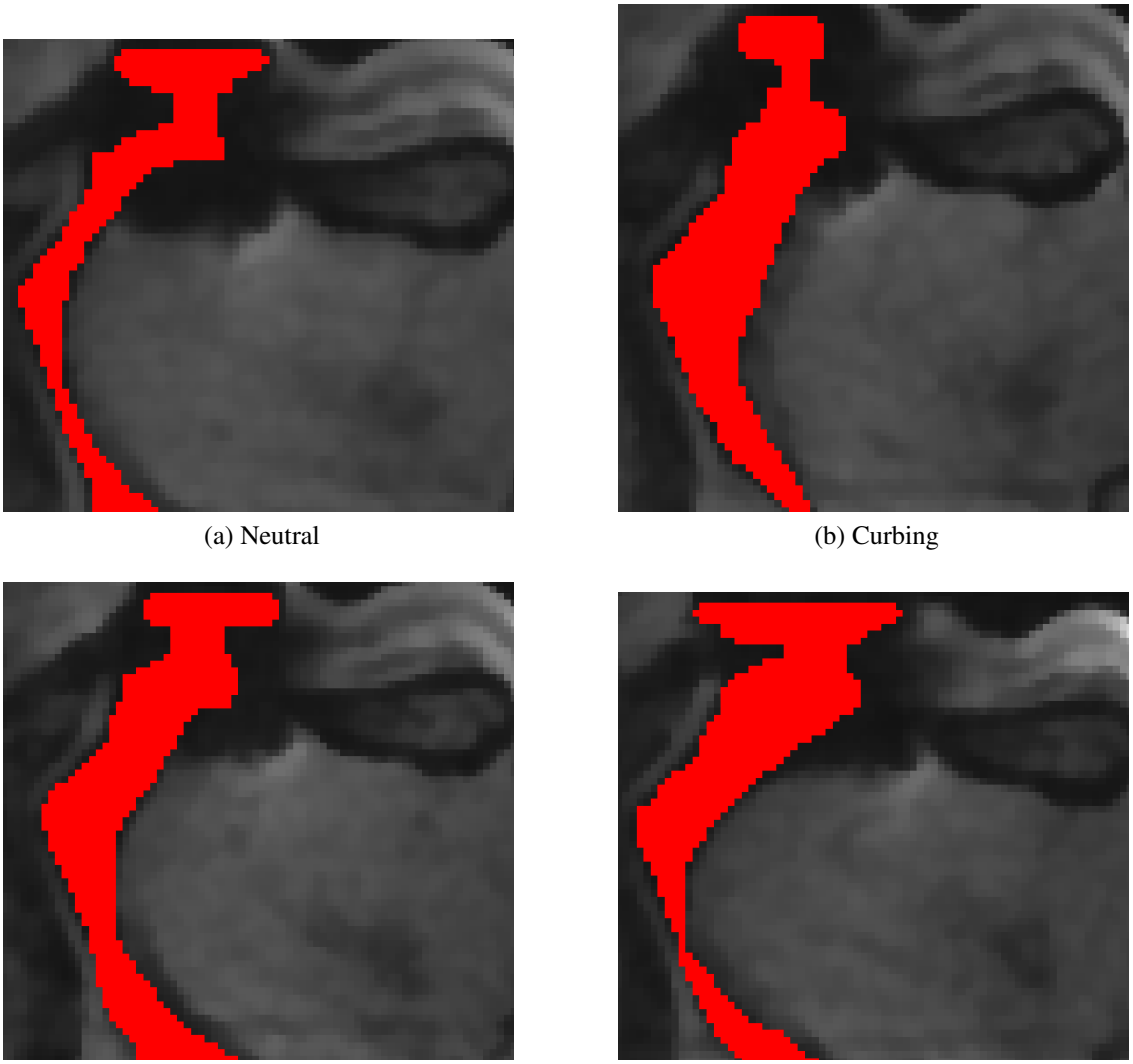
(a) Lip reconstruction



(b) Teeth reconstruction

Figure B.5: Screenshots of the reconstruction of lips and teeth shown on neutral

The outline of the teeth can be seen on some of the transverse slices, and from this the position is found where the front teeth are to be reconstructed. This still leaves the tip of the tongue which can only be seen on some of the photos, the tongue is reconstructed as seen in the saggital slices in figure B.6



(a) Neutral

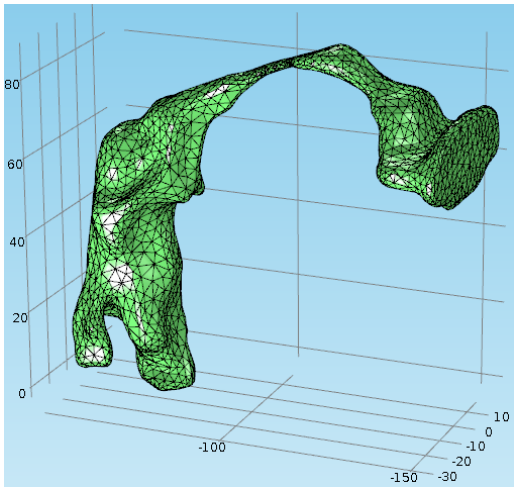
(b) Curbing

(c) Overdrive

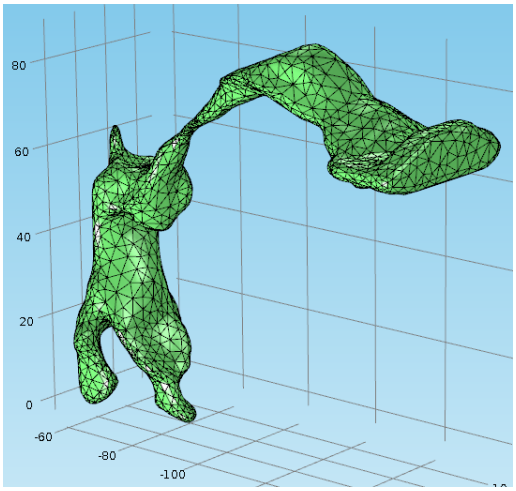
(d) Edge

Figure B.6: Reconstruction of the tongue in all modes

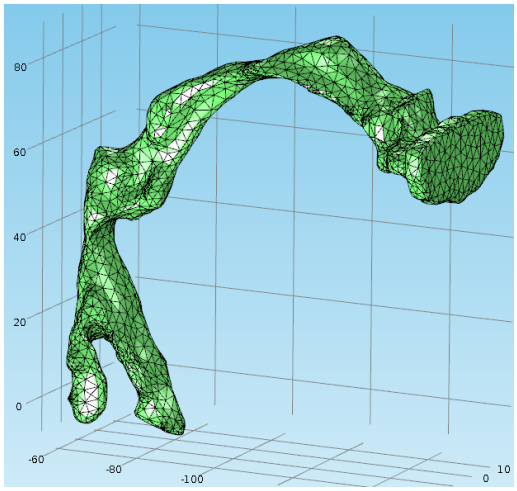
A slight smoothing of the three dimensional mask is performed using gaussian blur when it is fully drawn up. This simplifies the meshing by removing sharp edges. ScanIP creates a mesh of polygons on the surface of selected mask, which is exported to an STL file. A screenshot of the surface mesh of all modes can be found in figure B.7.



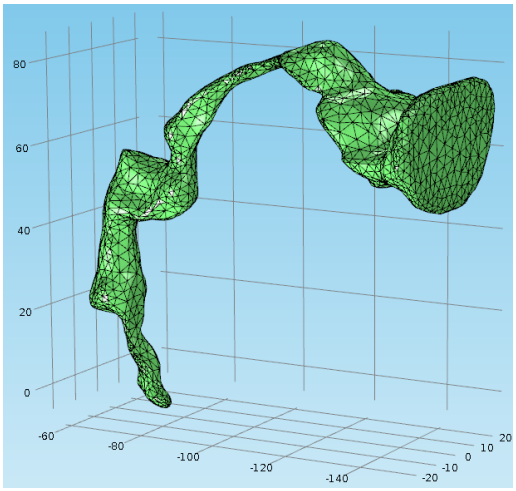
(a) Neutral



(b) Curbing



(c) Overdrive



(d) Edge

Figure B.7: The surface mesh of the segmented airway of the four modes

Note that versions of the surface mesh without the piriform sinuses are also produced, except for in Edge, where the piriform sinuses are closed.

APPENDIX C

Extraction of area functions

This analysis is performed in MATLAB, where the STL file is imported using a script called STL2MATLAB, where the 3d mesh without the piriform sinuses is used. This script creates a cell with three matrices, each containing rows of x, y and z coordinates of points in space. For each column the three matrices map out a triangle polygon. Another script was used to make the slices through these polygons called SLICE_ISO_DATA. This script uses two points in space to define the slice plane, a point on the plane and a point mapping out the normal vector. The script then identifies the polygon edges that intersect this plane, and creates a point at the intersection. The output of the script is a matrix of points in three dimensional space. These coordinates are then flattened into two dimensions for the area extraction.

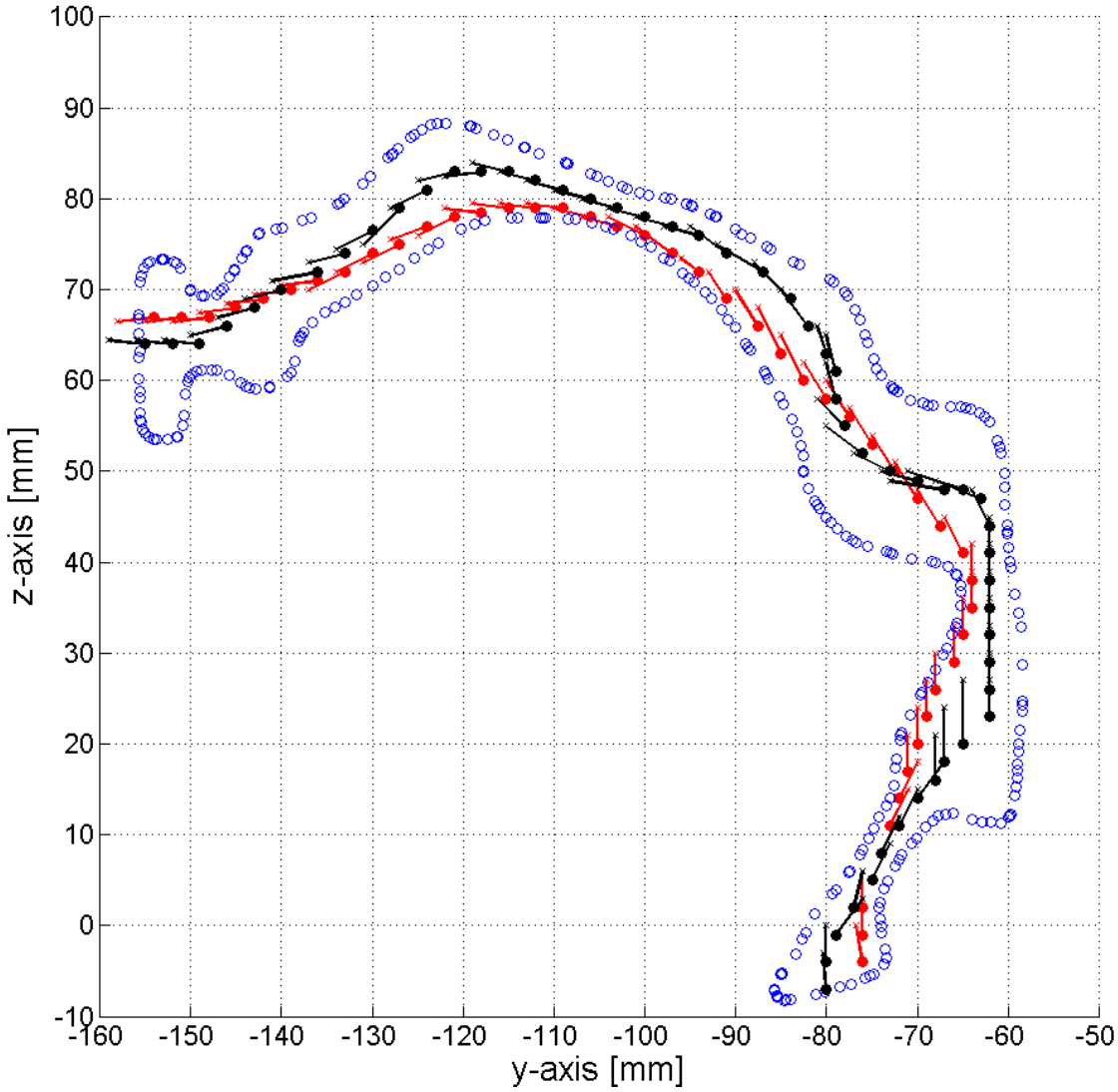


Figure C.1: Example of two slice selections. This figure shows a slice in the yz-plane through series no 4 in the "neutral" mode

The positions of the slices are all chosen manually by inserting the point on the plane and its normal. For the positioning procedure the whole vocal tract is visualized on a slice through the yz-plane (saggital plane) and the trail of slices is build up on this. An example is found in figure C.1, where a long and a short trail through the vocal tract is shown.

The trail of slices is order from the glottis so they can be used to compute the x-axis in the area-function. The trail chosen through the vocal tract will reflect upon how the sound will propagate through the vocal tract. In [Milenkovic et al., 2010] it was shown that the dominant effect of the curvature of the vocal tract is a lengthening of the mid-line around the convex curvature. The most correct trail is found by comparing to the numerical simulations and adjusting the trail to the best fit. The angle of the slice plane is chosen so the slice generally shows the smallest cross-section of the airway. The vocal tract is with

good approximation symmetrical in the yz-plane, so the slice planes are not angled in this direction. The red dots represent the slice plane and the cyan dots represent the vector that maps out the normal of the plane.

When a slice is created using the SLICE_ISO_DATA script, it some times includes parts of the model that is not supposed to be included, f.ex. when slicing parts of oral cavity, parts of the larynx is included. A small script was written excludes selected regions from the slices, so only the enclosed airway is left for analysis.

In the area function the nasal passage is excluded to simplify the analysis. Most of the modes show a closed nasal passage, but especially in curbing there is an opening to the nasal passage. This is however neglected for simplicity in this analysis, as the curbing mode can be produced with a closed nasal passage [Sadolin, 2012].

The points circumfering the airway boundary is filled with a polygon using another script called POLYFY. This script indexes the points so that the FILL and POLYAREA commands can process the points as a polygon. An example of the flattening of the slice and the polygon filling is shown figure C.2.

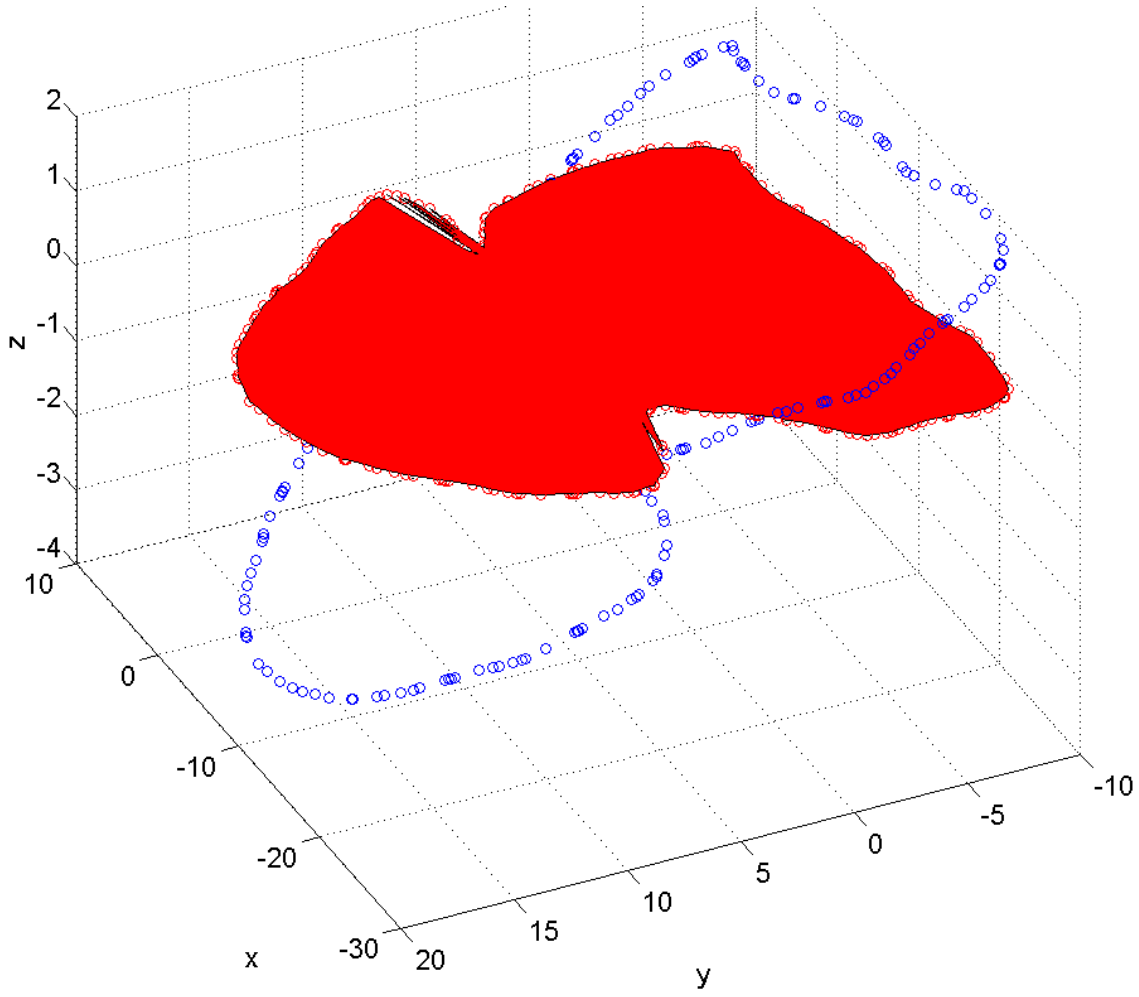


Figure C.2: Example of slice area extraction, slice no. 9 in series no. 4 in "neutral" showing the slice of where the larynx opens up into pharyngal cavity

The small errors in the polygon filling seen on the figure appears in some slices, but when they are small like the ones shown in the figure they are ignored. The gaps happen because the POLYFY script cannot identify the edge. All slices were checked and points were manually added or removed if the gaps in the polygon was too big.

APPENDIX D

VTMI description

A method of experimentally measuring the impedance of the vocal tract is used as another link in the verification of the extracted area functions. The Vocal Tract Measured Impedance method (VTMI) works by exciting the vocal tract with a known volume velocity and measuring the pressure response at the mouth opening. It is described in Epps et al. [1997] and Kob and Neuschaefer-Rube [2002], but the measurement method used here is inspired also by a method to measure the greens function in a room [Jacbosen, 2011]. The setup for the VTMI measurement is shown in figure D.1.

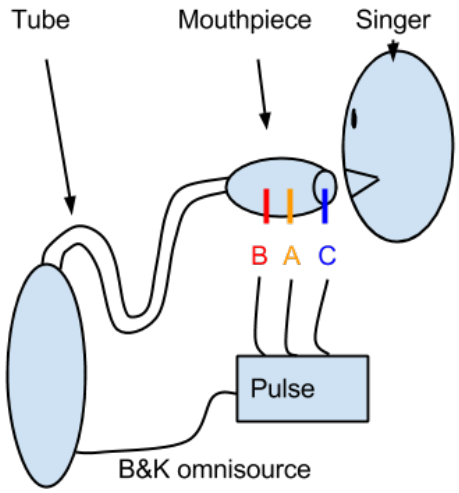


Figure D.1: Illustration of the measurement setup of the VTMI method.

As seen in figure D.1, the B&K omnisource with a mouthpiece is used as a source. The mouthpiece is equipped with 2 pressure microphones, which makes it possible to measure the volume velocity of the source through finite difference approximation of the pressure gradient between the t wo microphones. This assumes that the sound field in the tube is one dimensional, which is a good approximation in the frequency range between 100 Hz and 3200 Hz. The microphones in the mouthpiece is denoted A and B as in the figure,

which means that the finite difference will be [Jacbosen, 2011]:

$$Q = \frac{S}{\rho c} \cdot \frac{p_A \cos(kl) - p_B \cos(k(l + \Delta l)}{j \sin k\Delta l} \tag{D.1}$$

where l is the distance between microphone A and the opening of the tube, and ΔL is the distance between the two microphones. The greens function describes the pressure response (microphone C) of the volume acceleration of a monopole at a given point in space, which in this case is right in front of the lips of the singer.

$$G(r, r_0) = \frac{p_C(x, y, z)}{j\omega\rho Q} \tag{D.2}$$

To find the impedance, which is the pressure over volume velocity. The greens function is simply scaled by $j\omega$ to integrate the volume acceleration to the volume velocity in the frequency domain.

By measuring a frequency response between two microphones it is possible to eliminate the noise that are in both microphones. The greens function can be rewritten as an expression that uses only the frequency responses between the 3 microphones ($H_{AC}, H_{BC} and H_{AB}$)[Jacbosen, 2011].

$$G(r, r_0) = \tag{D.3}$$
$$\frac{\sin(k\Delta l)}{kS} \frac{H_{AC} \cos(kl) - |H_{AB}|^2 H_{BC} \cos(k(l + \Delta l))}{\cos^2(kl) - 2Re\{H_{AB}\} \cos(kl) \cos(k(l + \Delta l)) + |H_{AB}|^2 \cos^2(k(l + \Delta l))}$$

The remaining results

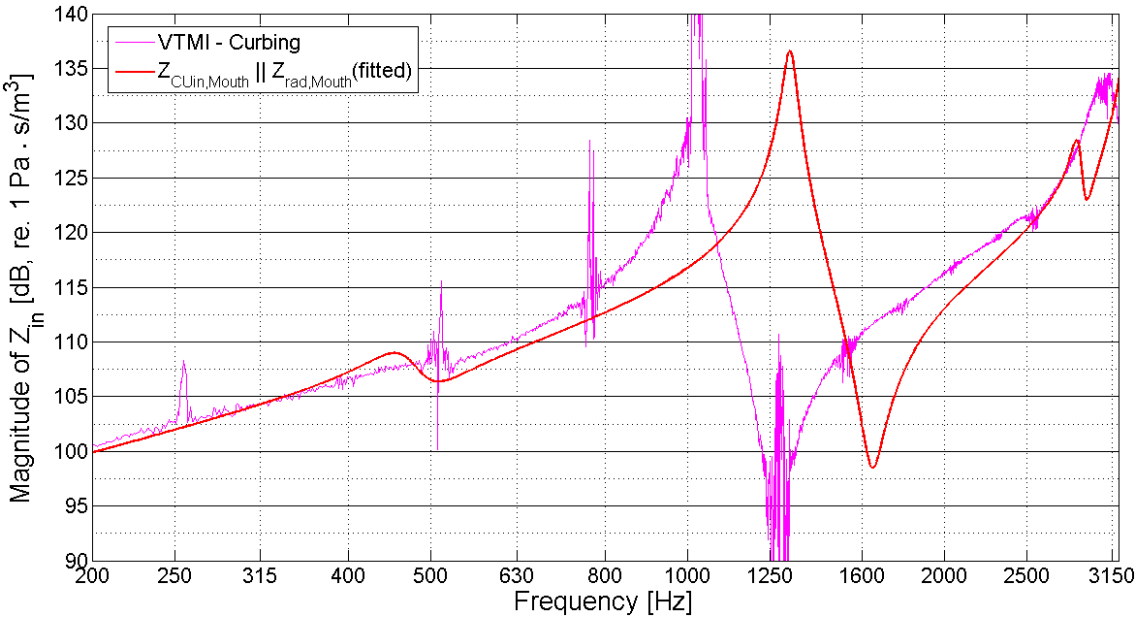


Figure D.2: Result of the VTMI measurement for Curbing, compared to a simulation of the measurement.

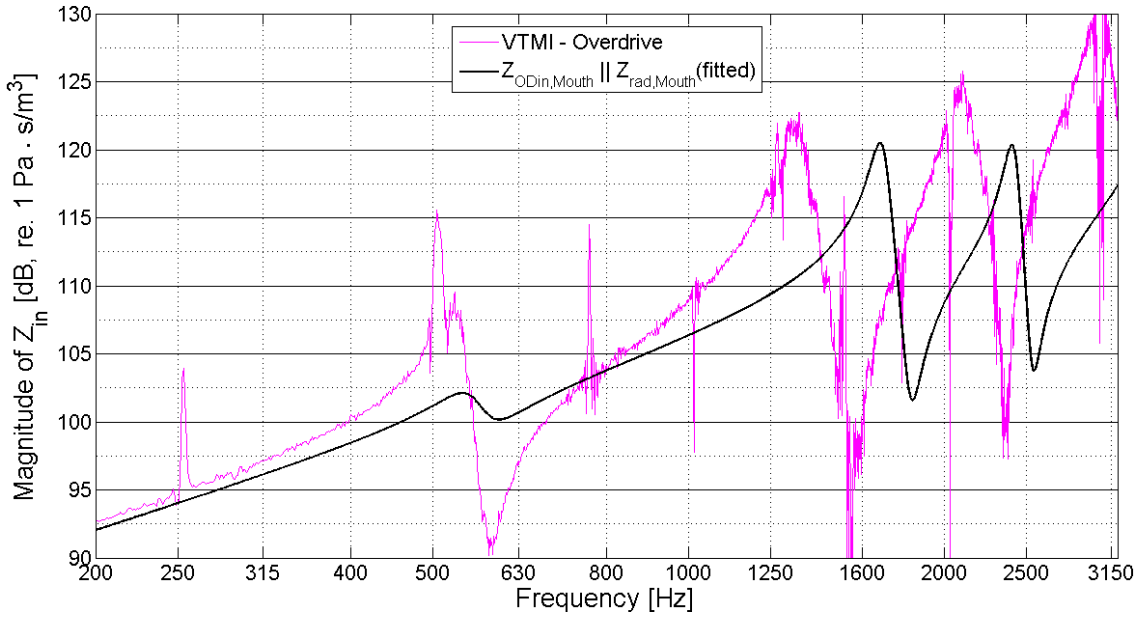


Figure D.3: Result of the VTMI measurement for Overdrive, compared to a simulation of the measurement.

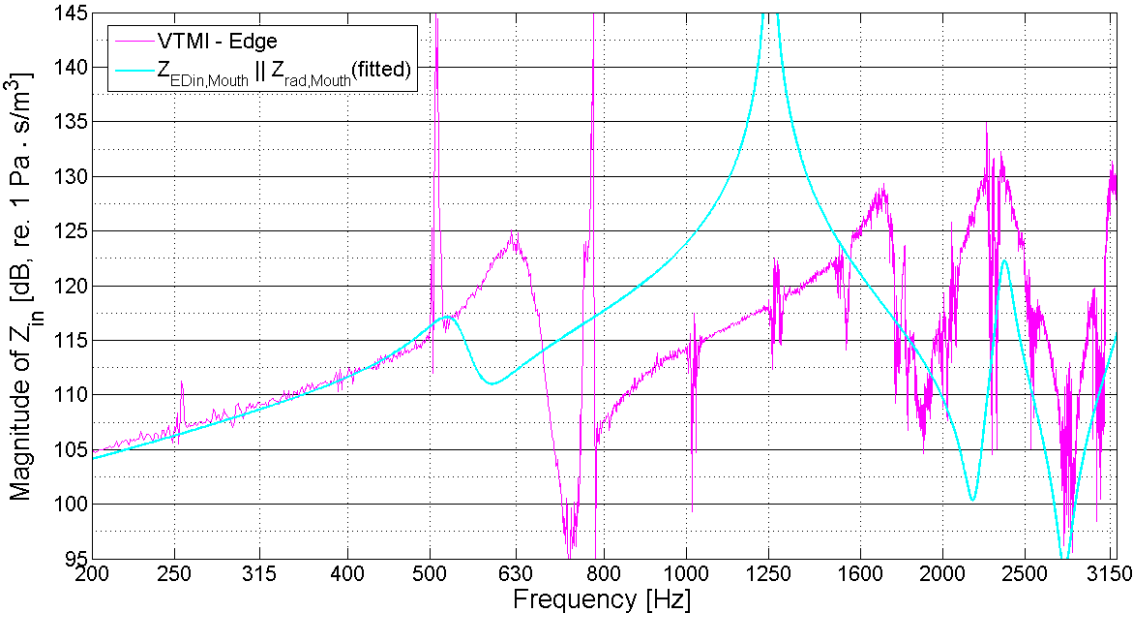


Figure D.4: Result of the VTMI measurement for Edge, compared to a simulation of the measurement.

Bibliography

Adachi, S. and Yu, J. [2005], 'Two-dimensional model of vocal fold vibration for sound synthesis of voice and soprano singing', *J. Acoust. Soc. Am.* **1**17(5), 3213–3224.

Austin, S. F. and Titze, I. R. [1997], 'The effect of subglottal resonance upon vocal fold vibration', *J Voice* **1**1(4), 391–402.

Backstrom, Alku and Vilkman [2002], 'Time-domain parameterization of the closing phase of glottal airflow waveform from voices over a large intensity range', *IEEE Transactions on Speech and Audio Processing* **1**0(3), 186–192.

Baken, R. J. [1992], 'Electroglottography', *JOURNAL OF VOICE* **6**(2), 98–110.

Birkholz, P., Kroger, B. and Neuschaefer-Rube, C. [2011], 'Synthesis of breathy, normal, and pressed phonation using a two-mass model with a triangular glottis', *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH* pp. 2681–2684.

Clement, P., Hans, S., Hartl, D. M., Maeda, S., Vaissiere, J. and Brasnu, D. [2007], 'Vocal tract area function for vowels using three-dimensional magnetic resonance imaging. A preliminary study', *J Voice* **2**1(5), 522–530.

COMSOL [2012], 'Comsol acoustics module user guide @ONLINE'.
URL: `http://hpc.mtech.edu/comsol/pdf/aco/AcousticsModuleUsersGuide.pdf`

Epps, J., Smith, J. R. and Wolfe, J. [1997], 'A novel instrument to measure acoustic resonances of the vocal tract during phonation', *Measurement Science and Technology* **8**(10), 1112–1121.

Farsight [2009], 'http://www.farsight-toolkit.org/ @ONLINE'.
URL: *http://www.farsight-toolkit.org/wiki/ITK_Pre-Processing_Algorithm_Wrappers_in_Python/GradientAnisotropicDiffusionFilter*

Flanagan and Landgraf [1968], 'Self-oscillating source for vocal-tract synthesizers', *IEEE Transactions on Audio and Electroacoustics* **1**6(1), 57–64.

Fulop, S. A. [2011], *Speech Spectrum Analysis*, Springer.

gbmc.org [2013], 'The voice center, greater baltimore medical center @ONLINE'.
URL: *http://www.gbmc.org/home_voicecenter.cfm?id=1552*

GetBodySmart.com [2013], 'Getbodysmart.com @ONLINE'.
URL: *http://www.GetBodySmart.com*

Hanson, L. G. [2009], 'Introduction to magnetic resonance imaging techniques @ONLINE'.
URL: *http://eprints.drcmr.dk/37/1/MRI_English_a4.pdf*

Heil, M. and Hazel, A. L. [2011], 'Fluid-structure interaction in internal physiological flows', *Annual Review of Fluid Mechanics* **4**3.

Henrich, N., d'Alessandro, C., Doval, B. and Castellengo, M. [2004], 'On the use of the derivative of electroglottographic signals for characterization of nonpathological phonation', *J. Acoust. Soc. Am.* **1**15(3), 1321–1332.

Henrich, N., Smith, J. and Wolfe, J. [2011], 'Vocal tract resonances in singing: Strategies used by sopranos, altos, tenors, and baritones', *J. Acoust. Soc. Am.* **1**29(2), 1024–1035.

Ishizaka, K. and Flanagan, J. L. [1972], 'Synthesis of voiced sounds from a two-mass model of the vocal cords', *The Bell System Technical Journal* **5**1(6), 1233–1268.

Jacbosen, F. [2011], 'Green's function in a room', *Exercise guide used in the course "Advanced Acoustics" at DTU spring 2012* .

Jacobsen, F. [2010], 'Fundamentals and acoustics of noise control', *Lecture note for couse by the same name* .

Kob, M. and Neuschaefer-Rube, C. [2002], 'A method for measurement of the vocal tract impedance at the mouth', *Med Eng Phys* **2**4(7-8), 467–471.

Liljencrants, J. [1985], 'Speech synthesis with a reflection-type line analog', *Dept. of Speech Communication and Music Acoustics, Royal Institute of Technology, Stockholm, Sweden,* .

Lous, Hofmans, Veldhuis and Hirschberg [1998], 'A symmetrical two-mass vocal-fold model coupled to vocal tract and trachea, with application to prosthesis design', *Acustica - Acta Acustica* **8**4(6), 1135–1150.

Man, S. and Schroeter [1987], 'A hybrid time-frequency domain articulatory speech synthesizer', *IEEE Transactions on Acoustics, Speech, and Signal Processing* **3**5(7), 955–967.

Milenkovic, P. H., Yaddanapudi, S., Vorperian, H. K. and Kent, R. D. [2010], 'Effects of a curved vocal tract with grid-generated tongue profile on low-order formants', *J. Acoust. Soc. Am.* **1**27(2), 1002–1013.

Pelorson, Hirschberg, Wijnands and Bailliet [1995], 'Description of the flow through in-vitro models of the glottis during phonation', *Acta Acustica* **3**(2), 191–202.

Phonetics Science, A. [2010], 'Praat: Sound to formant @ONLINE'.
URL: *http://www.fon.hum.uva.nl/praat/manual/Sound__To_Formant__burg____.html*

Rosen, C. B. and Simpson, C. B. [2008], *Operative techniques in larygology*, Springer.

Sadolin, C. [2012], *Komplet Sangteknik, 5.edition*, Shout Publications.

Sadolin, C. and McGlashan, J. [2007], 'Unpublished study of complete vocal technique and vocal effects @ONLINE'.
URL: *http://www.youtube.com/watch?v=nJu_BQrfk3E*

Selamtzis, A. [2011], 'Acoustics of the singing voice', *DTU* .

Sondhi, M. M. [1974], 'Model for wave propagation in a lossy vocal tract', *Journal of the Acoustical Society of America* **5**5(5), 1070–1075.

Steinecke, I. and Herzel, H. [1995], 'Bifurcations in an asymmetric vocal-fold model', *J. Acoust. Soc. Am.* **9**7(3), 1874–1884.

Story, B. [1995], *Physiologically-based Speech Simulation Using an Enhanced Wave-reflection Model of the Vocal Tract*, University of Iowa.
URL: *http://books.google.dk/books?id=8PsNywAACAAJ*

Story, B. H., Laukkanen, A. M. and Titze, I. R. [2000], 'Acoustic impedance of an artificially lengthened and constricted vocal tract', *J Voice* **1**4(4), 455–469.

Story, B. H. and Titze, I. R. [1995], 'Voice simulation with a body-cover model of the vocal folds', *J. Acoust. Soc. Am.* **9**7(2), 1249–1260.

Sundberg, J. [1987], *The science of the singing voice*, Northern Illinois University Press.
URL: *http://www.GetBodySmart.com*

Sundberg, J., Scherer, R., Hess, M., Muller, F. and Granqvist, S. [2013], 'Subglottal pressure oscillations accompanying phonation', *J Voice* **2**7(4), 411–421.

Sundberg, J. and Thalen, M. [2010], 'What is "Twang"?', *J Voice* **2**4(6), 654–660.

Titze, I. and Alipour, F. [2006], *The Myoelastic Aerodynamic Theory of Phonation*, National Center for Voice and Speech.
URL: *http://books.google.dk/books?id=TdhbBAAACAAJ*

Titze, I. R. [1984], 'Parameterization of the glottal area, glottal flow, and vocal fold contact area', *J. Acoust. Soc. Am.* **7**5(2), 570–580.

Titze, I. R. [1988], 'The physics of small-amplitude oscillation of the vocal folds', *J. Acoust. Soc. Am.* **8**3(4), 1536–1552.

Titze, I. R. [1989], 'A four-parameter model of the glottis and vocal fold contact area', *Speech Communication* **8**(3), 191 – 201.
URL: *http://www.sciencedirect.com/science/article/pii/0167639389900010*

Titze, I. R. [2002], 'Regulating glottal airflow in phonation: application of the maximum power transfer theorem to a low dimensional phonation model', *J. Acoust. Soc. Am.* **1**11(1 Pt 1), 367–376.

Titze, I. R. [2008], 'Nonlinear source-filter coupling in phonation: theory', *J. Acoust. Soc. Am.* **1**23(5), 2733–2749.

Titze, I. R. and Story, B. H. [2002], 'Rules for controlling low-dimensional vocal fold models with muscle activation', *J. Acoust. Soc. Am.* **1**12(3 Pt 1), 1064–1076.

Titze, I. R. and Strong, W. J. [1975], 'Normal modes in vocal cord tissues', *Journal of the Acoustical Society of America* **5**7(3), 736–744.

Titze, I. R. and Worley, A. S. [2009], 'Modeling source-filter interaction in belting and high-pitched operatic male singing', *J. Acoust. Soc. Am.* **1**26(3), 1530.

Wikipedia [2012], 'Body planes wiki @ONLINE'.
URL: `https://en.wikipedia.org/wiki/File:BodyPlanes.jpg`

List of Figures

List of Tables