Technical University of Denmark

Maria Chantziara (s131176)

Acoustic loading of the vocal folds of the singing voice

Master's Thesis, July 2015

DTU Electrical Engineering
Department of Electrical Engineering

MARIA CHANTZIARA (S131176)

Acoustic loading of the vocal folds of the singing voice

Master's Thesis, July 2015

Supervisors:

Finn T. Agerkvist

Alba Granados

DTU - Technical University of Denmark, Kgs. Lyngby - 2015

Acoustic loading of the vocal folds of the singing voice

This report was prepared by:
Maria Chantziara (s131176)

Advisors:
Finn T. Agerkvist
Alba Granados
Jonas Brunskog

Project period: January 2015- July 2015

ECTS: 35

Education: MSc

Field: Engineering Acoustics

Table of Contents

Abstract

Ongoing scientific research concerning the mechanism of the singing voice, is conducted the last years. The identification of the basic parameters that control the functionality of the vocal system, would give singers the possibility to produce desirable sounds with less effort and in a healthy way. Complete Vocal Technique is a singing method which is based on the anatomy and physiology of the vocal system. This technique suggests four vocal modes with which the human instrument can produce sounds, independently the singing style of the singer.

The shape of the vocal tract during the phonation of the modes and the strength of the harmonics was defined. The geometrical data were obtained from previous research using Magnetic Resonance scanning. Already recorded data were processed. The results showed that each vocal mode has a different epilarynx tube, which seems to influence the strength of the coupling between the source and the filter. The analysis of the recorded data, showed evidence that there are two possible modes between the *non metallic* and *full metallic* modes. The fact that different vowels were used for each mode, renders the conclusion preliminary and suggests further experimental approach. The vocal folds model is a symmetrical mechanical system. Each vocal fold consists of two masses, springs and dampers. The system self oscillates without the presence of the vocal tract. The obtained vocal tract configurations for four modes were coupled with the vocal folds model. The tension parameter was adjusted to shift the fundamental frequency of the phonation and the initial displacement of the masses triggered the oscillation. An interaction between the vocal folds and the vocal tract was observed since the strength of the harmonics was shifted when different modes were simulated. However, many of the harmonics seem underestimated compared to the measurements. The deviation from the measurements is attributed to the simplified model that was used. Unfortunately, the time restriction did not leave space for further analysis and the implementation of a more complex model.

Preface

This report is the documentation of a master thesis in Engineering Acoustics at the Technical University of Denmark (DTU). The project lasted from January 2015 to July 2015 under the supervision of Assoc. Prof. Finn Agerkvist. The study includes the modelling of the mechanics of the singing voice.

The motivation for the project was the ongoing research in the field of the singing voice and particularly the Complete Vocal Technique. Teaching techniques for singers are being improved and singers can understand in an easy way how to use their voice in a healthy way. Complete Vocal Technique is a singing method suggested by the Complete Vocal Institute and Cathrine Sadolin. The four vocal modes are introduced and analysed. Recent research on the modes is also discussed.

Special thanks to Cathrine Sadolin, Henrik Kjelin, Jesper Pedersen. Especially thanks to my supervisors, Finn Agerkvist and Alba Granados, for their advice, their patience and the useful directions they gave to me during this project.

Maria Chantziara (s131176)

CHAPTER 1

Introduction

The human voice is one of the most important means of communication. It is considered to have a twofold meaning since, apart from using it for producing speech, it is also used as an instrument for creating music. Theoretical models for the vocal fold oscillation and the wave propagation in the vocal tract are already suggested but an accurate representation of the phonatory system has yet to be developed. The anatomy and the physiology of the human phonatory system is, in general, well represented but the complexity of the functionality of the system demands further research.

The singing voice is under research the last years. Not only does it concern the scientists in order to create a realistic model of the singing voice, but it is important for pedagogical purposes that will help singing teachers to teach in a very effective way. Specifically, a realistic model would help to understand how the physical changes in the vocal tract and the vocal folds influence the produced sound.

Many studies have been conducted concerning the singing voice. In Sundberg [1994], the differences between the spoken vowels and the sung vowels were reviewed. Two aspects characterize the voiced sounds. The vowel quality and the voice quality. The vowel quality depends on the first two formants and give the vowel its characteristic sound. The voice quality depends mainly on the higher formants. In the latter study, it has been proved that in high pitches e.g. soprano singing, the vowel identification is very difficult because the first two formants are very close to each other.

Vowel articulation demands particular vocal tract shapes in every singing style. Varying the vocal tract, means that the resonances in the vocal tract are shifted due to the movement of the lips, the mouth and the jaw [Henrich et al., 2011]. In singing, when a harmonic of the voice is very close to a resonance of the vocal tract, this can increase

the sound level of the sound that is produced. This effect is known as *resonance tuning* and allows singers to perform efficiently by putting less effort. However, in some cases, it is suggested that in high pitches and especially in singing, the non linear coupling of the vocal tract and the vocal folds is more intense and can create frequency instabilities [Titze, 2008].

Complete Vocal Technique is a recent study which is based on the anatomy and physiology of voice production. It proposes four vocal modes of the singing voice a) Neutral, b) Curbing, c) Overdrive and d) Edge. The definition of the modes is based on subjective and objective measures, which prove the different features of each mode.

A very helpful tool to define the shape and the function areas of the vocal tract is the *Magnetic Resonance Imaging* or *MRI*. [Clement et al., 2007] conducted MRI measurements using a male test subject phonating spoken vowels. The results show that the MRI gives satisfactory results and can be used to represent the human vocal tract during phonation. The vocal tract of the four vocal modes have also been analysed by conducting MRI scans [Pedersen, 2013]. During the measurements the test subject was phonating different vowels performing the four vocal modes. The data from both MRI studies are used in the current project.

A thorough literature research was performed in order to familiarize with the anatomy and the physiology of the vocal system, and how this knowledge could be used in the modelling of the vocal tract and the vocal folds. The frequency and time domain model is used for the vocal tract analysis. For the vocal folds, the lamped parameter model and the continuous model are discussed but the latter is not used for further investigation due to the complexity of the system and limitations that were posed during the current study.

Great focus is given on the vocal tract shape. According to earlier research, the shape of the vocal tract can define the degree of interaction with the vocal folds [Austin and Titze, 1997]. The linear *source filter theory* suggests an independent behaviour of the vocal tract filter and the vocal folds source. Though, for high frequencies of the voice it is suggested that the non linear coupling dominates. This would explain an acoustic loading of the vocal folds and a possible change in their vibration [Titze, 2008].

The aim of the current study is to model the human vocal system and investigate the interaction between the vocal tract and vocal folds. The four vocal modes suggested by the *Complete Vocal Technique* will be used for the modelling.

The main parts of the theoretical background concerning the human phonatory system are presented in the second chapter. Special attention is paid to the source filter theory and the understanding of the four modes suggested by the Complete Vocal Technique. In the third chapter, the techniques for modelling the vocal tract and the vocal folds are analysed. The fourth chapter focuses on the four vocal modes and the vocal tract shapes each mode has. An analysis of the spectra of recorded data is presented. MRI data from previous studies are used for this purpose. In the fifth chapter, simulation results are analysed and discussed.

CHAPTER 2

Voice production

2.1 The human vocal system

The human speech consists of a sequence of sounds. The classification of these sounds is controlled by linguistic rules which are not part of this research since this is a subject of the linguistic sector. However, the knowledge of the structure of sounds is important for speech signal processing. Therefore, in order to analyse the speech signals a good understanding of the mechanism of their production is needed [Fakotakis, 2005].

Speech production is a complex mechanism and for this purpose many organs such as the throat, the nose, the mouth and a number of muscles are used. The speech production mechanism consists of several parts, some of which belong to the respiratory system and others to the digestive system. An illustration of the human phonatory system can be seen in Fig. 2.1 and it is composed by the lungs, the trachea, the larynx and the vocal tract.

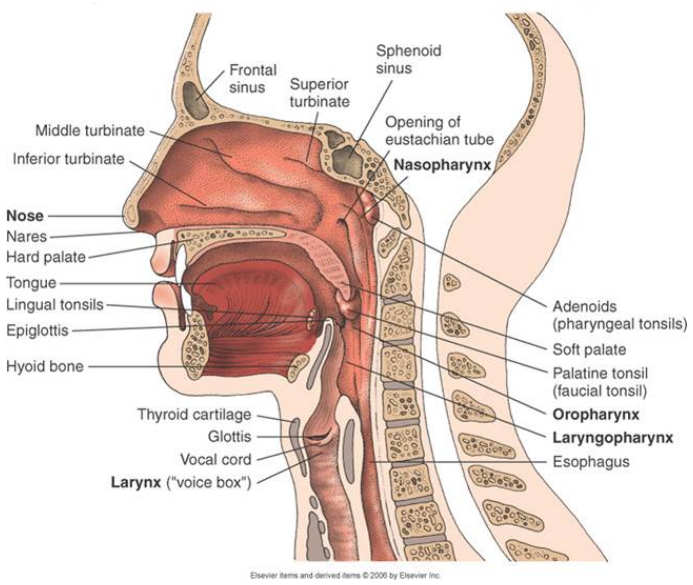


Figure 2.1: The human phonatory system [tdmu.edu, 2015].

Speech is produced when air is forced from the lungs up to the trachea past the vocal folds as they open and close in a periodic or quasi-periodic manner. The sound energy that is produced by the flow of air is defined as the glottal source. The frequency at which the vocal folds vibrate gives the sound its *pitch* and it is called the *fundamental frequency or* f_0. The fundamental frequency differs among individuals and it is within the range of $85-150\ Hz$ for a male, $165-255\ Hz$ for a female and $250-300\ Hz$ for a child.

The sound wave which is produced by the vocal folds is modified by the structures of the vocal tract. The vocal tract includes the throat, the nasal and mouth cavities. With the help of the pharynx, the soft palate, the lips, the jaw and the tongue the sound obtains its characteristic sound colour (or timbre) and by changing the position of each one of these structures, different vowels can be produced. When the vocal tract shape restricts or blocks the air, then the consonants are produced [Plack, 2005]. The consonants can be either voiced (the vocal folds vibrate) or unvoiced (the vocal folds do not vibrate). A block diagram of this procedure is depicted in Fig. 2.2.

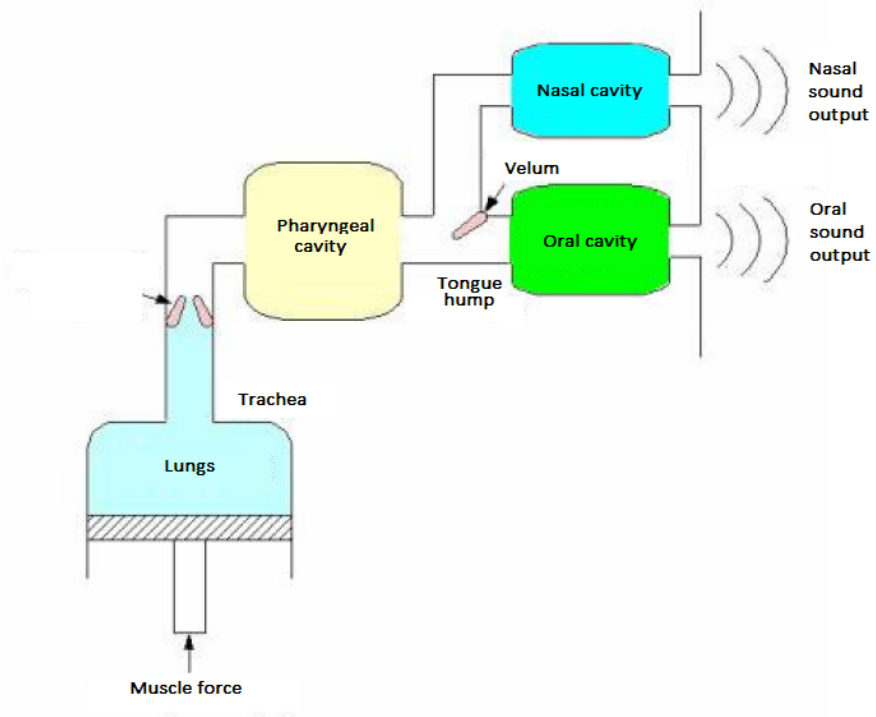


Figure 2.2: Speech production mechanism diagram [Fakotakis, 2005].

As mentioned above, the frequency at which the vocal folds vibrate is the fundamental frequency. However, the vocal signal is a complex signal which contains many simple sound waves which present a series of *harmonics* in the spectrum of the vocal signal. The harmonics are integer multiples of the fundamental frequency. For some frequencies, there are broad peaks in the spectra. These peaks are called formants and they result from the resonances of the vocal tract, see Fig. 2.7.

The larynx or the *voice box* contains the vocal folds and is composed by cartilages structured together, by ligaments and intrinsic (inside the larynx) and extrinsic (outside the larynx) muscles. At the upper part, the larynx is in direct continuity with the throat, hanging from the root of the tongue, which explains why its position is affected by the movements of the tongue. At the lower part, it lies on the top of the trachea, which continues to the lower respiratory tract. A frontal section of the larynx can be seen in Fig.2.3.

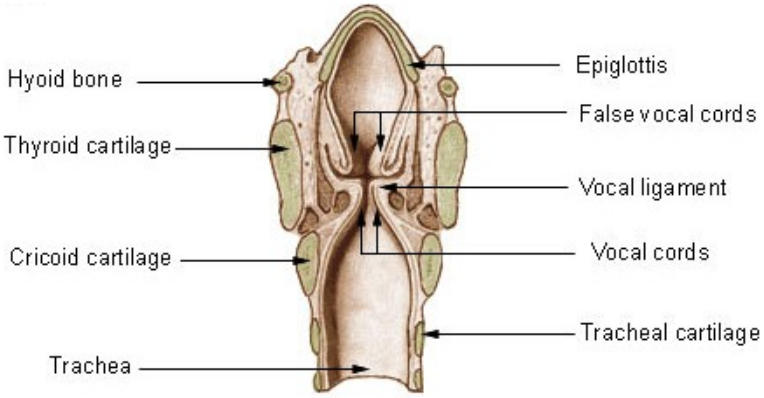


Figure 2.3: Frontal view of the larynx [Wikipedia, 2015].

The larynx contains nine cartilages, three single and three paired (Fig. 2.4). The largest laryngeal cartilage is the *thyroid cartilage* positioned in front of the vocal folds and serves to protect the larynx. Inferior the thyroid cartilage sits the *cricoid cartilage* which forms the back neck of the larynx(lamina). Along the upper edge of the cricoid lamina the *arytenoid cartilages* lie. They are small pyramid shaped paired cartilages and they are part of the larynx where the vocal folds are attached. They regulate the movement of the *vocalis muscles* which are responsible for the adduction and abduction of the vocal folds. Above the arytenoid cartilages, the *corniculate cartilages* are located and reinforce the arytenoids function. The *cuneiform cartilages* are paired cartilages which add support to the soft tissue. The epiglottis cartilage is attached to the back of the thyroid cartilage at the angle just below the superior thyroid notch and above the vocal folds. Its main role is to close the larynx during the swallowing process [GetBodySmart.com, 2015].

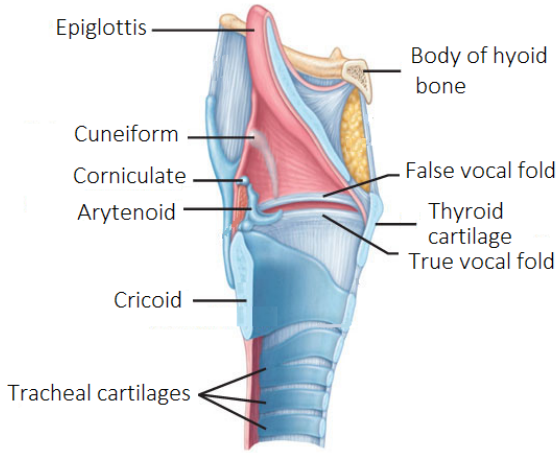


Figure 2.4: Lateral view of the cartilages of the larynx [StudyBlue.com, 2015].

The role of the cartilages is to hold together and support the six small intrinsic laryngeal muscles while the role of the intrinsic muscles is to control the tension and orientation of the ligaments that produce the voice [GetBodySmart.com, 2015]. More concretely, the

transversed arytenoid, the *lateral cricoarytenoid* and the *posterior cricoarytenoid* muscles help to bring the vocal folds together (adduction) and move them apart (abduction). The *vocalis*, the *thyroatenoids* and the *cricothyroids* change the tension of the vocal folds and consequently their length. Outside the larynx, there is a number of extrinsic muscles which support the larynx and raise or lower it within the trachea.

The vocal folds or vocal cords (true vocal folds) are two paired ligaments lined by a moveable mucous membrane and are controlled by muscles and arytenoid cartilages. In each vocal fold a vocal ligament is fixed in and when they vibrate they produce voiced sounds [Sadolin, 2012]. The vocal folds are multi layered structures. They consist of a deep layer, an intermediate and a superficial layer. The surface of the vocal folds is covered by a membranous tissue, the epithelium. The maximum length of the vocal folds is about $16mm$ for a male adult and $10mm$ for a female adult. The inner most dense part of the vocal folds is the thyroarytenoid muscle or vocalis. Above the vocal folds there are *vestibular* or *false vocal folds*. Each false vocal fold has a mucous ligament and a supportive vestibular ligament. An illustration of the multi layered structure of the vocal folds can be seen in Fig. 2.5.

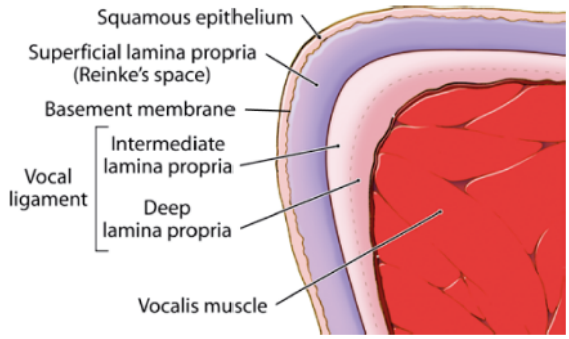


Figure 2.5: Frontal view of the vocal fold layers [Pedersen, 2013].

2.1.1 Vocal fold vibration

The vocal fold vibration is a sequence of multiple vibratory cycles. The vocal folds move due to the larynx muscles, the nerves and the cartilages. In order to understand how voice is produced, a single vibration of the vocal folds is described below in accordance with Fig. 2.6.

Glottal cycle

Early studies explained the vocal fold vibration using the *myoelastic aerodynamic theory*. Specifically, due to aerodynamic forces, the air pressure below the closed vocal folds

starts increasing. When this subglottal pressure overpowers the resistance of the vocal folds, the vocal folds start to blow open. The air continues to flow upwards and finally opens the top of the vocal folds. The elastic nature of the body makes the bottom of the vocal folds to start closing. This induces in decrease in pressure due to the *Bernoulli effect*. The adduction of the vocal folds produces a pulse of air [NCVS.org, 2015]. Then, the cycle begins again.

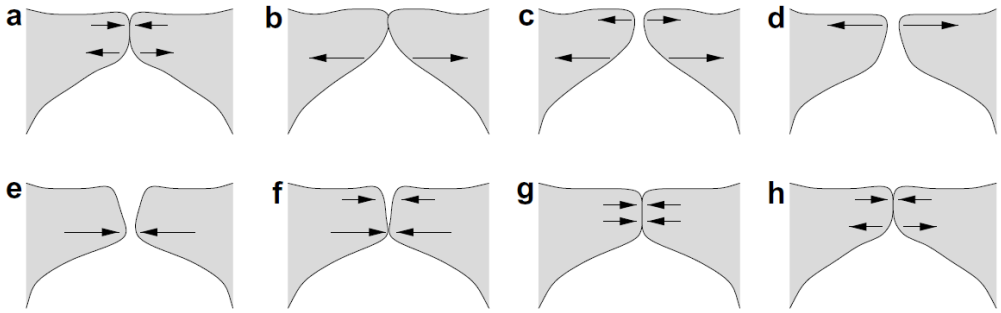


Figure 2.6: Vibration pattern of the vocal folds [Selamtzis, 2011].

The Bernoulli principle which is described above and is responsible for the vocal fold oscillation is defined as follows:

$$P + \frac{1}{2}\rho u^2 = \text{constant} \tag{2.1}$$

where P is the total pressure, ρ is the air density and u is the velocity. According to Eq. 2.1, an increase in the fluid velocity reduces the pressure between the folds so as their sum is constant. When the vocal folds are about to collide the pressure is very low. As mentioned above, during phonation the pressure below the closed glottis is higher than the pressure above the glottis and consequently the vocal folds open and the air starts to flow. However, the elastic forces of the vocal folds will try to close them again with the aid of the aerodynamic forces. Though, there is a region in which the Bernoulli effect is not present during the oscillation. This occurs when the flow separates from the boundary, which is at the glottal exit. At his case, the pressure and the volume velocity are constant. Thus, the jet effect is present [Pedersen, 2013].

2.1.2 Source filter theory

Speech production in humans is considered to stem from an interaction of a source of glottal flow, represented by the vocal folds, modulated by a filter which is represented as an acoustic resonator and is called the vocal tract. The acoustic characteristics of vowel and voiced consonant production are described by this model and is referred to as the *Source filter theory*. This theory has generally been described in literature [Fant, 1960] [Ishizaka

and Flanagan, 1972]. It is a linear source filter theory which assumes that the source of sound is independent of the filter. Concretely, in frequency domain the linear source filter interaction can be characterized by the multiplication of the Fourier transforms of the source signal and the filter transfer function, while in the time domain this interaction can be characterized by the convolution of the source and the impulse response of the filter [Titze, 2008]. The aforementioned source filter interaction is illustrated in Fig. 2.7.

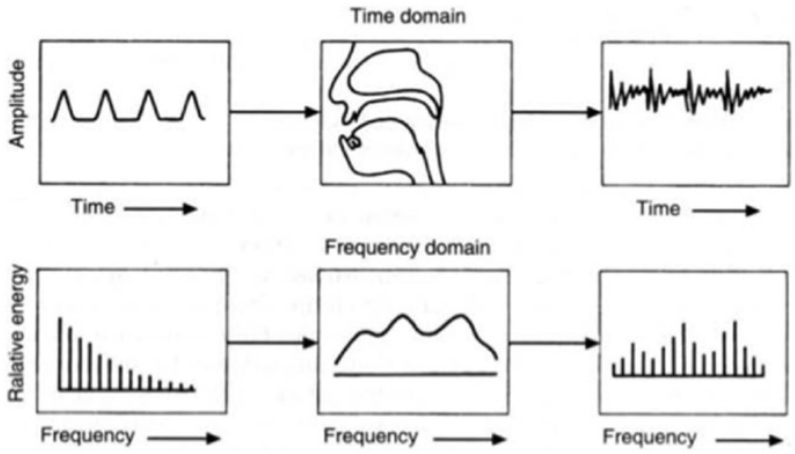


Figure 2.7: Sound production in speech [Chang, 2012].

The linear source filter theory describes adequately many aspects of speech production but only under specific conditions and simplifications. However, according to Titze [2008] the linear source filter theory describes better male speech rather than female or child speech. In male speech the harmonics are well below the formants of the vocal tract which implies a weak interaction between the source and the filter. But regarding female speech or singing, the harmonics can have higher values than the formants. This implies an intense interaction between the dynamics of the vocal folds vibration and the acoustic energy of the vocal tract. During this interaction, the acoustic airway pressure is responsible for the production of frequencies at the source. In extreme cases, the non-linear source filter coupling can result in source instabilities and produce sudden jumps of the fundamental frequency, subharmonics generation or chaotic vocal fold vibration [Titze and Worley, 2009] [Titze, 2008]. These results are attributed to the acoustic loading by the vocal tract and is not taken into account by the linear source filter theory.

When singers change the length of the vocal tract by elevating or lowering the larynx they consequently adjust the resonances or formants of the vocal tract. According to [Henrich et al., 2011], if a harmonic of the vocal signal has a value close to a formant the strength of this harmonic will increase. Moreover, in high frequencies the reactive load of the vocal tract is higher, especially when the fundamental frequency is very close to a resonance of the vocal tract [Titze, 1988] [Titze, 2008]. In particular, this formant and harmonic

tuning helps the singers to sing in a very high level while putting little effort. In order to achieve the desirable resonance tuning there should be a proper combination between the range of the resonance and a harmonic [Henrich et al., 2011]. The way the singers tend to combine formants and harmonics depends on the singing style and it is suggested that certain vowels and singing styles fit better with certain pitches [Titze, 2008].

2.2 Complete Vocal Technique

Complete Vocal Technique or CVT is a singing method which aims to offer a good understanding of voice production and to help singers to perform in a healthy way. According to Sadolin [2012] the technique is based on three fundamental principles that every singer should adopt. These are:

- support: to learn how to use the diaphragm and the muscles of the body.
- necessary twang: the area below the vocal folds (epiglottic funnel) is narrowed while the arytenoid cartilages approach the lower part of epiglottis.
- avoid protruding the jaw and tightening the lips: in this way unwanted constriction is prevented and singers achieve an easy transition going from consonants to vowels.

The concept of the CVT is to divide the vocal sounds in groups according to how metallic a sound is, independently of the musical style each singer has. In particular, four vocal modes are suggested, see Table 2.1.

Table 2.1: The four modes and the amount of how metallic they are [Sadolin, 2012].

Name of mode	Amount of metallic
Neutral	non metallic
Curbing	half metallic
Overdrive	full metallic
Edge	full metallic

The term metallic is a subjective measure and can be considered as a characteristic of a high pitch, bright, intense and direct sound. Further, the presence of metallic in a sound is usually combined with *twang*. Twang describes the moment when the the vocal tract becomes very narrow in the epilaryngeal and pharyngeal area [Titze, 2002]. [Hanayama et al., 2009] suggested that when singers perform a metallic sound they adjust their vocal

tract in a way that they raise the larynx, lower the velar and tighten the aryepiglottic funnel. Moreover, a significant shift of the second formant was observed in high frequencies and the increase in the intensity of the sound is attributed to the increase of the amplitude of the third and fourth formant [Hanayama et al., 2009].

Neutral

Neutral is a non metallic mode and can be identified by soft notes with or without air added. It can be performed by both males and females while singing all the vowels. The sound level of the neutral mode is, in general, low and quiet but it can be loud in high pitches i.e. above $C2(65.4\,Hz)$ for females and above $C1(32.7\,Hz)$ for males. The method of singing the mode in a correct way is to loose the jaw and keep the upper jaw in front of the lower one.

The vocal tract shape in neutral is characterized by a wide epilarynx tube. When the vocal tract is wide, the degree of interaction of the vocal tract and the vocal folds is low. This can justify the wide frequency range of vowels that can be phonated by both males and females Titze [2008] Pedersen [2013].

Curbing

Curbing is the softest metallic mode and can be recognized by a holding in the breath before a note is sung. It reminds the sound of a moan or cry and it has a medium volume. It can be sung only for specific vowels including I (as in sit), UH (as in hungry) and O (as in woman). As in neutral, curbing is eligible for both males and females to sing in the whole frequency range.

The vocal tract in curbing shows a similar shape as in neutral but it is narrower. Further, according to electroglottograph recordings (EGG), the duration of the vocal fold adduction in curbing is the lowest compared to the other modes which implies the metallic and the tight collision of the vocal folds Pedersen [2013].

Overdrive

Overdrive belongs to the metallic modes and is identified by its loudness and it resembles a shout. In the high part of the voice the mode is restricted in two vowels, EH (as in stay) and OH (as in so). In particular, males can sing up to $C2(65.4\,Hz)$ and females up to $D2(73.41\,Hz)$. The basic method that is followed to phonate in overdrive is to use the so called "bite" technique which demands to drop the jaw while the mouth is in smiling position, the lower jaw should be behind the upper one and the distance between the jaws is one finger length Sadolin [2012].

In overdrive the vocal tract is very narrow which implies strong interaction between the vocal tract and the vocal folds. In [Titze and Worley, 2009] it is described that if the input impedance of the vocal tract is comparable to the impedance of the source then the energy that is stored in the vocal tract due to the reactance can reinforce the vocal fold vibration through feedback (non-linear source filter interaction). This could explain the loudness of the mode and the restriction in the vowel articulation.

Edge

Edge belongs to the full metallic modes. It is characterized by a high sound pressure level, reminds a scream, and it can be sang by both males and females in all frequencies but only in twanged vowels. Specifically, the vowels that can be articulated are I (as in sit), EH (as in stay), A (as in and) and OE (as in herb).

As in overdrive, edge is characterized by a narrowed epilarynx tube which reinforces the acoustic loading of the source. Twang increases this interaction since it implies a narrowing of the epiglottic funnel and consequently a narrow vocal tract.

Even though the four vocal modes can be recognised subjectively, several studies including EGG and larynx endoscopy prove that there are significant differences among the modes. Electroglottography is a non invasive technique that permits the inspection of the vocal fold contact area during phonation [Henrich et al., 2004]. In order to conduct an EGG two electrodes are attached on the skin of the test subject on the thyroid cartilage of the larynx and a high frequency modulated current is applied [Henrich et al., 2004] [completevocalinstitute.com, 2015]. The main parts of the EGG signal, as seen in Fig.2.8, are when the vocal folds a) start closing (gradual increase in amplitude), b) are fully adducted and no air passes through (maximum amplitude), c) start opening (gradual decrease in amplitude) and d) are fully abducted (flat signal).

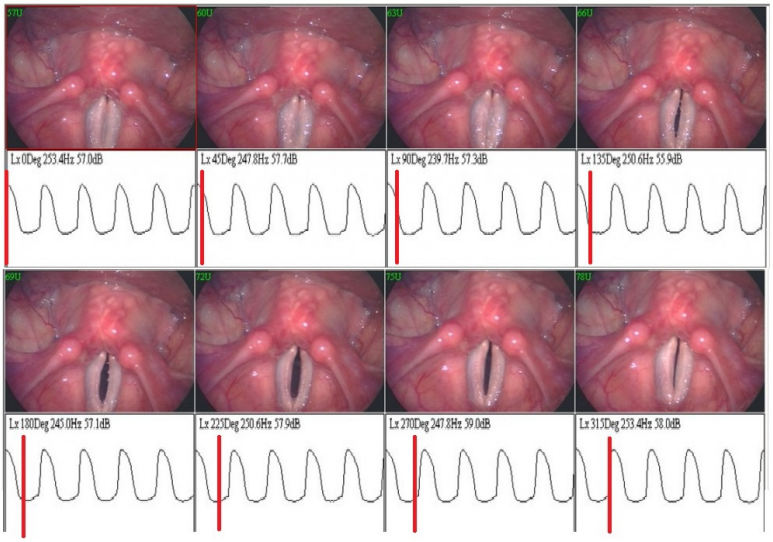


Figure 2.8: Synchronised larynx stroboscopy with EGG [completevocalinstitute.com, 2015].

Using the results of the EGG measurements the *contact quotient* Q_x, which describes how long the vocal folds remain adducted, can be estimated. Q_x is calculated by measuring the width of the waveform at a point 70% of the height from the peak of the waveform of the EGG signal divided by the period of the signal [completevocalinstitute.com, 2015].

Recently, the four modes are under discussion and especially the curbing mode. Between neutral and the two full metallic modes, overdrive and edge, two modes are introduced: *half edge* and *half overdrive*. Every sound that does not match with the aforementioned modes is considered to be in the curbing area. Two other modes that have been identified through recordings and belong to the curbing area are *medium edge* and *medium overdrive*.

In the current study the four modes (neutral, curbing, overdrive and edge) will be discussed and analysed. In addition, the four modes will be compared with the half overdrive, half edge, medium overdrive and medium edge based on recordings from the Complete Vocal Institute.

CHAPTER 3

Computational models

The focus of this chapter lies on the different concepts and techniques for modelling the vocal tract and the vocal folds.

For the vocal tract modelling, a frequency and a time domain model is discussed. The vocal tract consists of coupled tubes. The tubes have the same length but differ in cross sectional area (see Fig. 3.1). Both models make use of the one-dimensional wave equation. In frequency domain, it is assumed that in each junction the output pressure and velocity of the proceeding tube is the input pressure and velocity of the following tube. In time domain, the *wave reflection analog* is implemented. The latter model makes use of the reflection coefficients at the junctions of the tubes in order to calculate the pressure or flow that is passes or is reflected at each junction at each time step [Story, 1995].

For the vocal folds modelling two methods are generally suggested. The lumped element model and the continuum model. The lumped parameter model represents the vocal folds by using mechanical oscillators, springs and dampers. In this chapter, a two mass model will be described. Even though the two mass model is a simple model, it gives satisfactory representation of the interaction between the airflow and tissue movement to produce oscillation [Alipour et al., 2000]. However, for a more realistic realistic representation of the vocal folds continuum models have been developed. With the continuum models, the modelling of the multi layered structure of the vocal folds is more realistic. For the solution of the equations of motion, the finite element method or other numerical analysis is used. An introduction in the continuum model is presented at the end of this chapter. Due to time limitations, the model is not used for further analysis. The difficulty for using the continuum model lies in the complexity of the model. For instance, detailed geometrical data and the internal stress are needed to model the structure of the vocal folds. Moreover, the vocal folds constitute of three layers and five parameters for each of them are needed to be defined, as described in Section 3.2.2. These parameters are unknown and their values are based on assumptions that would lead in uncertainties.

3.1 The vocal tract

The vocal tract is an acoustical lossy non uniform tube. It is terminated by the lips and the nose which give the vocal tract the favourable shape to produce sounds, e.g. vowels and voiced consonants. The length of the vocal tract is approx. $17\,cm$.

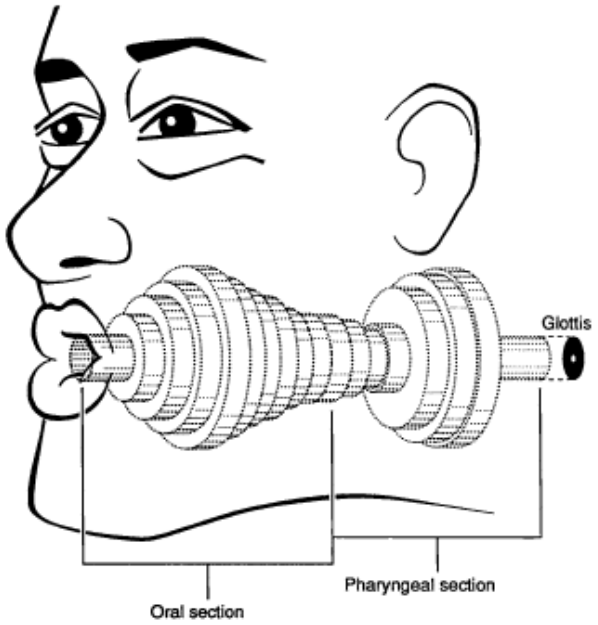


Figure 3.1: Representation of the vocal tract constituting of coupled tubes [Story, 1995].

For the analysis, in the next section, two approximations are made. Firstly, it is assumed that there is no bending at the back of the tongue and the vocal tract is simplified to a tube with a straight axis [Story, 1995]. Therefore, the one dimensional wave equation can be implemented [Flanagan et al., 2008]. However, the uniform tube configuration is unlikely for a human vocal tract because it demands the same width of both the larynx and the pharynx [Titze and Worley, 2009]. Second, it is considered that the largest cross sectional area of the vocal tract should be much smaller than a wavelength. This happens for frequencies no higher than $5\,kHz$. The limiting frequency is given by the following formula as mentioned in Flanagan et al. [2008]:

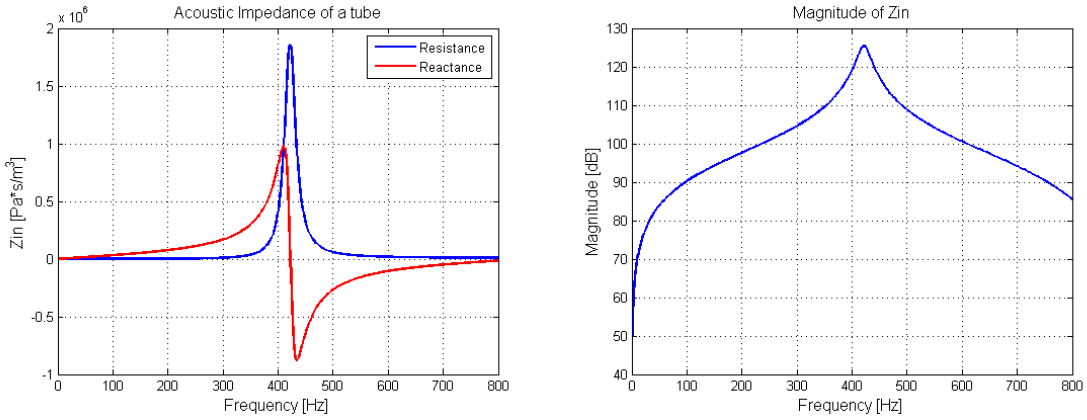
$$f_c = \frac{0.5861c}{2\alpha} \tag{3.1}$$

where c is the speed of sound and α is the radius of the cross sectional area [Story, 1995].

3.1.1 Frequency domain model of the vocal tract

A first approach of the vocal tract modelling was made by using the uniform tube as a simple vocal tract shape. A metric that describes the acoustic properties of a tube is the

acoustic impedance. The acoustic impedance is defined as the complex ratio of the sound pressure over the volume velocity. The input acoustic impedance of the vocal tract affects both the shaping of the glottal flow pulse and the vibrational pattern of the vocal folds [Story et al., 2000]. The real part of the acoustic impedance is the resistance while the imaginary part is the reactance. When the reactance is positive, the system behaves as a mass of air which is accelerated and decelerated. The acoustic loading of the latter system is characterized as inertive. When the reactance is negative, the behaviour of the system is springlike and the acoustic loading is characterised as compliant. For the case of zero reactance the load is resistive. The inertive load indicates the acoustic energy that is stored in the system as kinetic energy while the compliant load indicates the potential energy that is stored [Pedersen, 2013] [Titze, 1988].



(a) (b)

Figure 3.2: Input acoustic impedance of a uniform tube (a) Reactance and resistance (b) Magnitude.

In Fig. 3.2 the input acoustic impedance of a uniform tube is illustrated. The length of the tube is $17.6\,cm$ and the cross sectional area $S = 50\,cm^2$.

In Flanagan et al. [2008] it is mentioned that the radiation on a piston in a spherical baffle is a function of frequency and the relative size of the piston and the sphere. When the radius of the piston becomes smaller than that of the sphere the radiation can be approximated as the radiation of a piston in an infinite baffle. The radiation impedance of a piston mounted in an infinite baffle is given by the following formula [Jacbosen, 2011]:

$$Z_r = \frac{\rho c}{S}\left[\left(1 - \frac{J_1(2k\alpha)}{k\alpha}\right) + \left(\frac{K_1(2k\alpha}{2(k\alpha)^2}\right)\right] \tag{3.2}$$

where ρ is the density of air, S is the cross sectional area of the tube, k is the wavenumber, α is the radius of the piston, $J_1(x)$ is the first order Bessel function and $K_1(x)$ is a related

Bessel function [Flanagan et al., 2008]. For small values of $k\alpha$, i.e. $k\alpha < 0.5$, Eq.3.2 can be approximated as:

$$Z_r = \frac{\rho c}{S}\left(\frac{1}{2}(k\alpha)^2 + j\frac{8}{3\pi}k\alpha\right) \tag{3.3}$$

where k is the wavenumber and α is the radius of the cross sectional area at the end of the tube.

As seen in Fig. 3.2 the impedance is inertive (positive) till the first resonance. After the first resonance the load becomes compliant (negative).

In order to understand how the shape of the vocal tract influences the formants, a simulation of three different shapes of the vocal tract was conducted. The first shape (Fig. 3.3) is a uniform tube with a cross sectional area of $4\,cm^2$. For the second (Fig. 3.4) and the third shape (Fig. 3.5) a two tube vocal tract was used representing an idealized shape of the vowels $/a/$ and $/i/$ respectively. The cross sectional areas of the small and the big tube are $0.5\,cm^2$ and $3\,cm^2$ respectively. For all the three configurations a zero radiation impedance at the end of the tube was considered. As it can be seen from the graphs, the

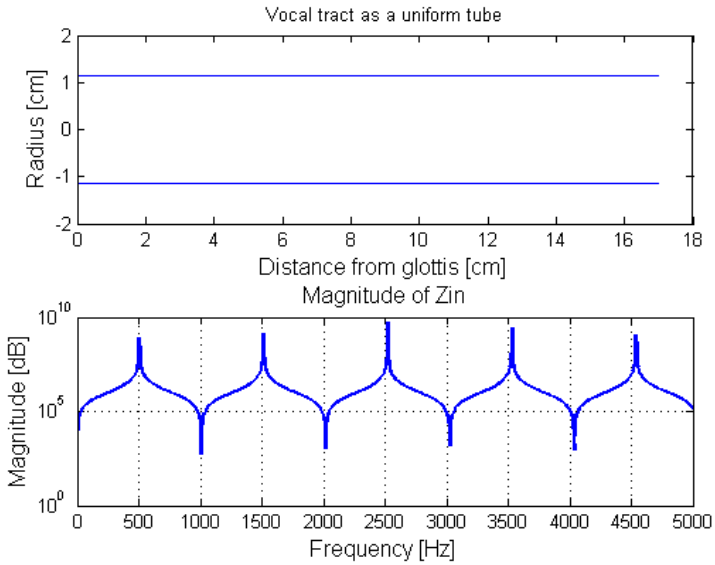


Figure 3.3: Vocal tract shape and magnitude of the acoustic input impedance of the uniform tube.

uniform tube, closed at one end, presents resonances at odd multiples of the quarter of the wavelength. This can be written as follows:

$$f_{res} = (2k+1)\frac{c}{4l}, \text{k=0,1,2,...} \tag{3.4}$$

where $c = 343\,m/s$ is the speed of sound and $l = 17\,cm$ is the length of the tube. The first four vocal tract resonances are at $504.4\,Hz, 1513\,Hz, 2522\,Hz, 3530\,Hz$. When the

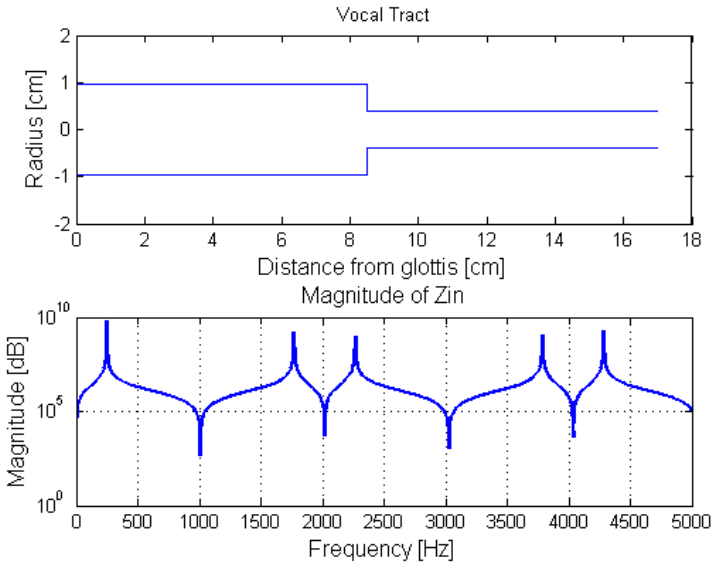


Figure 3.4: Idealized vocal tract shape and magnitude of the acoustic input impedance for the vowel $/i/$.

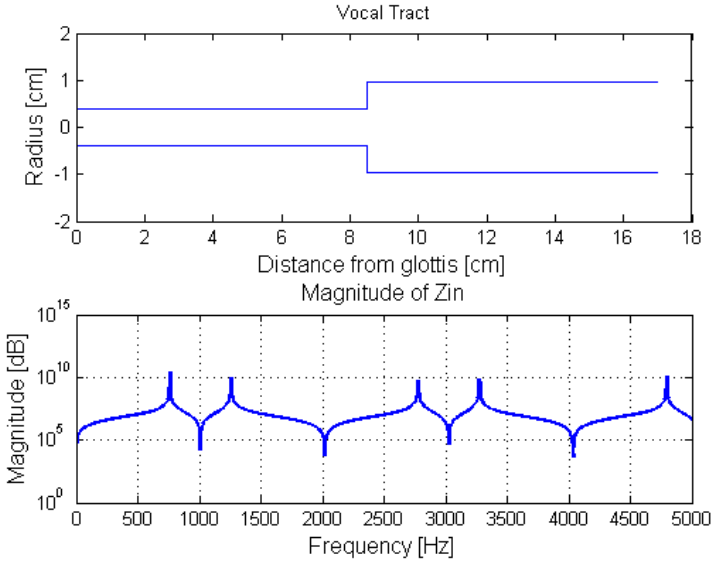


Figure 3.5: Idealized vocal tract shape and magnitude of the acoustic input impedance for the vowel $/a/$.

front tube is constricted (Fig. 3.4) the first and third formant move downward and the second and fourth upward. When the back tube is constricted (Fig. 3.5) the first and the third formant move upward while the second and the fourth move downward.

Fig. 3.4 and Fig. 3.5 represent systems of coupled tubes. At low frequencies, the sound field of each component can be considered one dimensional. Thus, at each junction the output sound pressure and volume velocity of the preceding component equals the input sound pressure and volume velocity of the following component. Consequently, in a two tube system the input acoustic impedance of the first tube is the acoustic load of the second tube. Let us assume a n tubes vocal tract. According to Jacbosen [2011], such a system can be described by its transmission matrix giving a relationship between the pressure and volume velocity at the first tube (input) and the sound pressure and volume velocity of the nth tube (output). This system is described by the following formula:

$$\begin{bmatrix} P_{in} \\ U_{in} \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} P_{out} \\ U_{out} \end{bmatrix} \tag{3.5}$$

where $P_{in}, P_{out}, U_{in}, U_{out}$ are the input and output pressures and volume velocities respectively. The transmission matrix is obtained by the multiplication of n transmission matrices calculated for every tube:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} A_1 & B_1 \\ C_1 & D_1 \end{bmatrix} \begin{bmatrix} A_2 & B_2 \\ C_2 & D_2 \end{bmatrix} ... \begin{bmatrix} A_n & B_n \\ C_n & D_n \end{bmatrix} \tag{3.6}$$

The parameters A_n, B_n, C_n, D_n for each of the n tubes are defined as:

$$\begin{aligned} A_n &= cos(kl_n) \qquad & B_n &= j\frac{\rho c}{S_n} sin(kl_n) \\ C_n &= j\frac{S_n}{\rho c} sin(kl_n) & D_n &= cos(kl_n) \end{aligned} \tag{3.7}$$

The input acoustic impedance is defined as the ratio of the input sound pressure of the input volume velocity. Using the transmission line parameters the input impedance of the tube is given by the following formula:

$$Z_{in} = \frac{AZ_{rad} + B}{CZ_{rad} + D} \tag{3.8}$$

where Z_{rad} is the radiation impedance at the end of the nth. It can be written as:

$$Z_{rad} = \frac{j\omega RL}{R + j\omega L} \tag{3.9}$$

where

$$R = \frac{128 Z_{mouth}}{9\pi^2}, \quad L = \frac{8\sqrt{\frac{S_{mouth}}{\pi}} Z_{mouth}}{3\pi c} \tag{3.10}$$

and $Z_{mouth} = \dfrac{\rho c}{S_{mouth}}$ is the characteristic impedance of the last tube of the vocal tract which represents the mouth, with a cross sectional area S_{mouth}.

3.1.2 Time domain: Wave reflection model

In the time domain, the *wave reflection model* is implemented which is described by Story [1995]. The vocal tract model is considered to be a sequence of n coupled tubes. In the current study the number of the coupled tubes was $n = 44$. It is assumed that the time needed for the sound wave to propagate exactly the distance from one tube to the next one is one time step as seen in Fig. 3.6 (adapted from [Pedersen, 2013]).

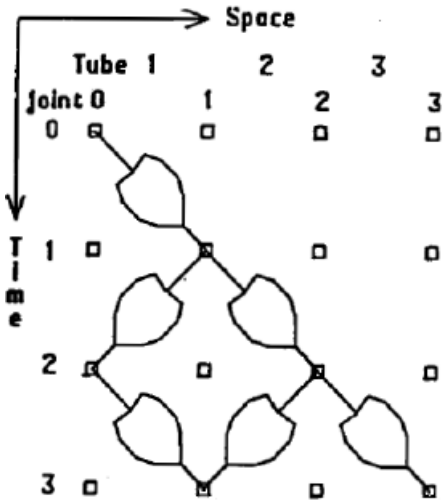


Figure 3.6: Wave propagation in the vocal tract [Pedersen, 2013].

For each time step, the reflection coefficient and the scattering is calculated at every junction. For the simple case of a two tube junction, and according to the solution of the one dimensional wave equation, the pressure and volume velocity of a tube can be considered as a forward going and backward going wave system (Fig. 3.7).

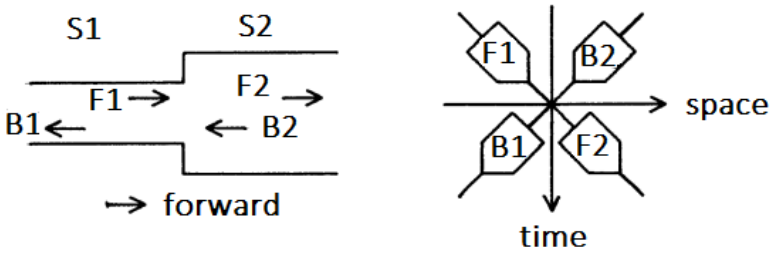


Figure 3.7: Two tube junction (left) and space time diagram of partial waves (right) [Liljencrants, 1985].

In fig. 3.7, S_1 and S_2 are the cross sectional areas of the two tubes. F_1 and F_2 represent the forward wave and, B_1 and B_2 the reflected (backward) wave. At every junction the total pressure is the sum of the forward and the reflected pressure. The total volume velocity is the difference of the forward and backward volume velocities, u_n^+ and u_n^- respectively.

The total pressure and the total volume velocity can be written as follows:

$$P_n = F_n + B_n \tag{3.11}$$

$$U_n = u_n^+ - u_n^- \tag{3.12}$$

But, the characteristic impedance of a tube is the ratio of the total pressure over the total volume velocity. Thus, the volume velocity at the nth section of the tube can be defined as the ratio of the difference of the partial pressures over the characteristic impedance $Z_n = \frac{\rho c}{S_n}$ of the respective tube. Eq. 3.12 becomes:

$$U_n = \frac{1}{Z_n}(F_n - B_n) \tag{3.13}$$

The wave reflection model uses the reflection coefficients in order to to calculate the wave propagation in each junction. The reflection coefficients are calculated by solving the one dimensional wave equation. At the junction of two consecutive tubes the reflection coefficients for each pressure wave (forward and reflected) are defined as:

$$r_{n,1} = \frac{S_n - S_{n+1}}{S_n + S_{n+1}} \tag{3.14}$$

$$r_{n,2} = \frac{S_{n+1} - S_n}{S_n + S_{n+1}} \tag{3.15}$$

where S_n is the cross sectional area of the nth tube. Let us assume that the pressures and velocities in the junctions are continuous. This can be written as follows:

$$F_n + B_n = F_{n+1} + B_{n+1} \tag{3.16}$$

$$\frac{F_n - B_n}{S_n} = \frac{F_{n+1} - B_{n+1}}{S_{n+1}} \tag{3.17}$$

After solving the system of Eq. 3.16 and Eq. 3.17, the forward pressure F_{n+1} and the reflected pressures B_n are given by the following scattering equations:

$$B_n = F_n \cdot r_n + B_{n+1} \cdot (1 + r_{n,2}) \tag{3.18}$$

$$F_{n+1} = F_n \cdot (1 + r_{n,1}) + B_{n+1} \cdot r_{n,2} \tag{3.19}$$

In Eq. 3.18 and 3.19 the pressures B_n and F_{n+1} are calculated using the pressures calculated one time step before. For a better understanding of the time step and the wave propagation in the coupled tubes, part of the Matlab code used for the simulations is apposed in Appendix A. The propagation of F_n and B_n is due to an excitation pulse at the

first section of the vocal tract [Pedersen, 2013]. This pulse represents the flow through the glottis and is defined as follows

$$F_1 = U\frac{\rho c}{S_1} \tag{3.20}$$

The radiation impedance at the end of the the vocal tract, as explained in the frequency domain, is given by Eq.3.3 which is the radiation impedance of the mouth. The reflection coefficient at the mouth can be written as:

$$r_{mouth} = \frac{Z_{rad} - Z_{mouth}}{Z_{rad} + Z_{mouth}} \tag{3.21}$$

Though, in order to use the radiation impedance Z_{rad} in Eg.3.9 the term $j\omega$ (see Eq.3.9) should be replaced and be brought into the time domain. This is achieved by implementing the mathematical transformation (bilinear transform) described in Story [1995]:

$$j\omega = \frac{2(1 - z^{-1})}{T(1 + z^{-1})} \tag{3.22}$$

where T is the sample interval and z^{-1} is a sample delay. By substituting the aforementioned transformation in Eq.3.21, the radiation impedance is calculated in the time domain. This radiation impedance is used for the calculation of the reflection coefficient at the mouth. The scattering equations are calculated according to Story [1995].

In Fig. 3.8 the results from the simulations using the frequency domain model and the wave reflection model are illustrated.

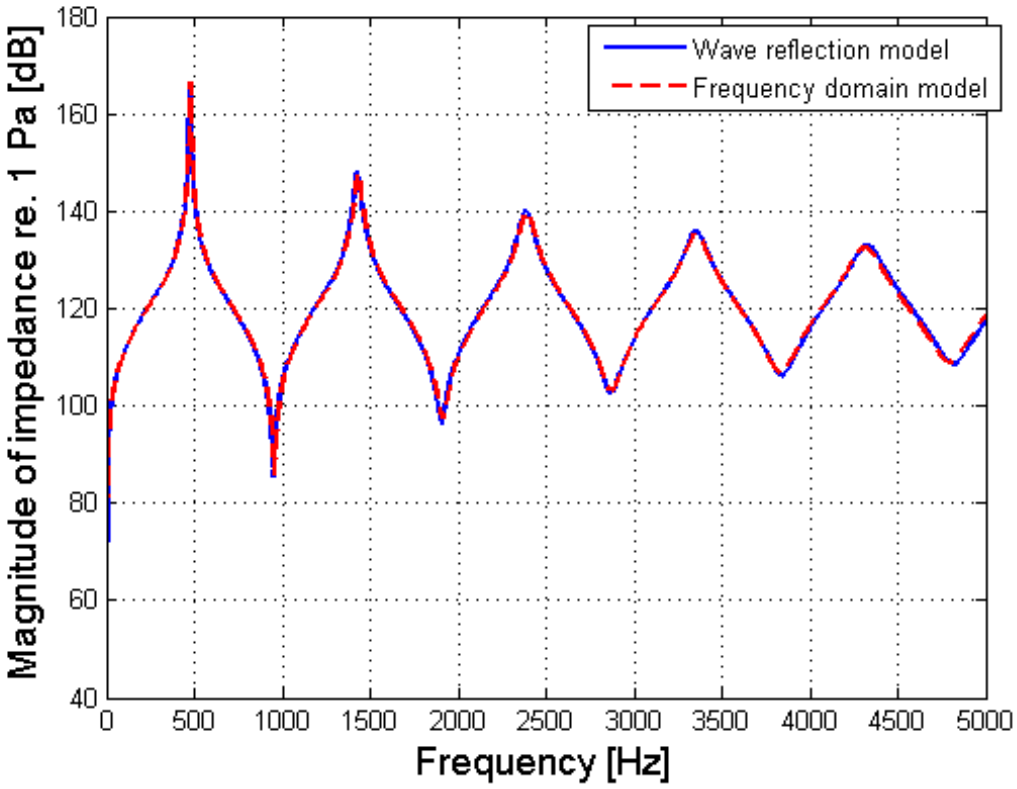


Figure 3.8: Wave reflection model and frequency domain model.

The frequency domain input acoustic impedance results from the calculation of the fast Fourier transform of the logarithmic amplitude of the input acoustic impedance of the wave reflection model. According to Fig. 3.8 this is verified and suggests the proper function of both models.

3.2 The vocal folds

This section is devoted to the vocal fold modelling. In Section 3.2.1, the lumped parameter model is discussed in detail. The vocal folds are represented as mechanical oscillators based on a combination of suggested theoretical models [Ishizaka and Flanagan, 1972][Steinecke and Herzel, 1995][Birkholz et al., 2011]. The theoretical background of the continuous model is discussed in Section 3.2.2 and is based on the finite element model developed by Alipour et al. [2000].

3.2.1 Lumped parameter model

The simplified two mass model studied in this project consists of two coupled mechanical oscillators of masses m_1 and m_2 which approximate the vocal folds. The masses are

placed vertically above one another [Steinecke and Herzel, 1995] [Titze, 1976]. The elastic properties of the vocal folds are represented by the springs k_1 and k_2, and by the dampers r_1 and r_2. The displacements x_1 and x_2 are the time-varying positions of the masses and represent the tissue deformations. An illustration of the system can be seen in Fig. 3.9.

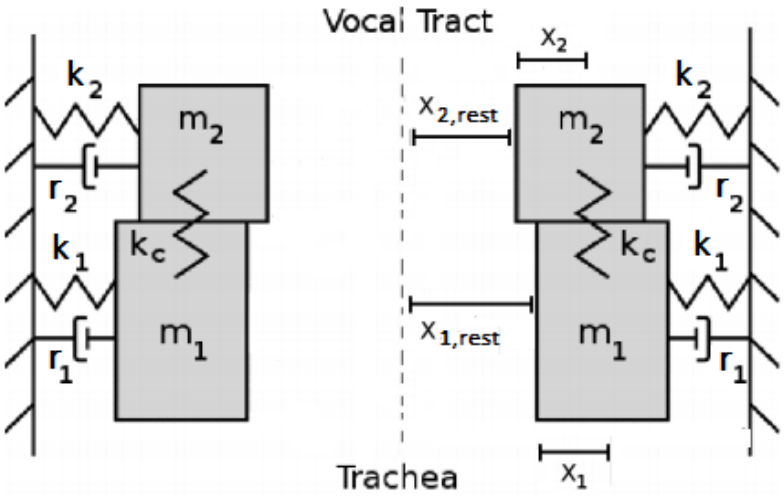


Figure 3.9: Two mass model of the vocal folds (Assaneo, 2013).

The main principle of the model is the phase difference between the upper and the lower part of the vocal folds. In order to produce oscillation the surface wave propagation in the mucosa must be from bottom to top. Thus, in the two mass system the bottom mass must always lead the top mass in phase [Titze, 1988].

The assumptions that were considered during the modelling include a symmetric version of the model, absence of non linear properties of the tissues and independence between the vocal fold vibration and the vocal tract resonances. The vocal fold vibration is flow induced [Steinecke and Herzel, 1995][Titze, 1988].

The equations of motion for each of the two masses are described as follows (adopted from Steinecke and Herzel [1995]):

$$F_{e1} = m_1\ddot{x}_1 + r_1\dot{x}_1 + kx_1 + k_c\left((x_1 - x_{1,rest}) - (x_2 - x_{2,rest})\right) + F_{1col} \tag{3.23}$$

$$F_{e2} = m_2\ddot{x}_2 + r_1\dot{x}_2 + kx_2 - k_c\left((x_1 - x_{1,rest}) - (x_2 - x_{2,rest})\right) + F_{2col} \tag{3.24}$$

where F_{e1} and F_{e2} are aerodynamic forces. F_{1col} and F_{2col} are forces that model the collision of the vocal folds and are defined as:

$$F_{1col} = \Theta(-\alpha_1)c_1\left(\frac{\alpha_1}{2l_g}\right) \tag{3.25}$$

$$F_{2col} = \Theta(-\alpha_2)c_2\left(\frac{\alpha_2}{2l_g}\right) \tag{3.26}$$

where α_1 and α_2 represent the lower and upper glottal areas and l_g is the length of the glottis. $\Theta(x)$ is a function that models the activation of the collision of the masses.

$$\Theta(x) = \begin{cases} 1, & \text{if } x > 0 \\ 0, & \text{if } x \leq 0 \end{cases} \tag{3.27}$$

The opening of the glottal areas between the lower and the upper masses are described as:

$$\begin{aligned} \alpha_1 &= 2l_g(x_1 + x_{1,rest}) \\ \alpha_2 &= 2l_g(x_2 + x_{2,rest}) \end{aligned} \tag{3.28}$$

The variables and the parameters that were used in the model were adopted from Steinecke and Herzel [1995], Birkholz et al. [2011] and Pedersen [2013]. They are presented in Table 3.1.

Table 3.1: Description and values of the variables used in the two mass model (adapted from [Pedersen, 2013].

Variable	Description	Value	Unit
m_1	mass 1	0.125	g
m_2	mass 2	0.025	g
r_1	damping constant 1	0.02	Ns/m
r_2	damping constant 2	0.02	Ns/m
k_1	spring constant 1	80	N/m
k_2	spring constant 2	8	N/m
k_c	coupling constant	25	N/m
d_1	thickness 1	2.5	mm
d_2	thickness 2	0.5	mm
c_1	collision spring constant 1	$3k_1$	N/m
c_2	collision spring constant 2	$3k_1$	N/m

The equations of motion mentioned above are connected to the aerodynamic forces through the glottal area [Story, 1995]. The oscillation of the system occurs due to the intraglottal pressure on the upper and the lower mass. The intraglottal pressure depends on the opening area of the glottis. According to Story [1995], the flow is diverging when there is minimum glottal area. The Bernoulli flow is present only below the narrowest part of the glottis [Steinecke and Herzel, 1995]. Above this part, the jet flow is formed and a

constant pressure is assumed. During the jet, the pressure is constant and is defined as follows:

$$P = P_i \tag{3.29}$$

where P_i is the supraglottal pressure. The upper mass is always subjected to P_i. When the Bernoulli effect is present, the pressure is given by the following equation:

$$P = P_s \left(1 - \Theta(a_{min}) \left(\frac{a_{min}}{a_1}\right)^2\right) \Theta(a_1) \tag{3.30}$$

The lower mass is subjected to the above mentioned pressure and P_s is the subglottal pressure. The volume velocity is defined as:

$$U = \sqrt{\frac{2P_s}{\rho}} a_{min} \Theta(a_{min}) \tag{3.31}$$

where a_{min} is the minimum area between the two masses. Additional forces F_{e1} and F_{e2} are acting on the two masses. These are aerodynamic forces which act in the intraglottal area. They are given by the following formulas:

$$F_{e1} = l d_1 P_1 \tag{3.32}$$

$$F_{e2} = l d_2 P_2 \tag{3.33}$$

where P_1 and P_1 are the pressures exerted at the lower and upper mass respectively, l is length of the vocal folds, d_1 and d_2 the thickness of the mass 1 and mass 2 respectively.

Vocal tract and vocal folds coupling

The coupling of the glottal flow and the wave propagation in the vocal tract is described by Titze [1984]. Two travelling waves act upon the glottis, one subglottal, with pressure P_s and one supraglottal with pressure P_i. The aforementioned pressures are obtained from calculations, using the wave reflection model of the vocal tract. Flow continuity is assumed between the subglottal and supraglottal areas, A_s and A_e respectively. The source filter interaction is presented using the following formula

$$u_g = \frac{a_g c}{k_t} \left\{ -\frac{a_g}{A^*} \pm \left[\left(\frac{a_g}{A^*}\right)^2 + \frac{4k_t}{pc^2} * (r_s P_s - r_e P_i \right]^{\frac{1}{2}} \right\} \tag{3.34}$$

Eq.3.34 is the analytical solution for the glottal flow where u_g is the glottal flow, a_g is the time varying glottal area, c is the speed of sound and k_t is a transglottal pressure coefficient for modified Bernoulli flow through the glottis [Titze, 1984][Titze, 2008]. A^* is an effective vocal tract area defined as

$$A^* = \frac{A_s \cdot A_e}{A_s + A_e} \tag{3.35}$$

The reflection coefficients r_s and r_e are given by the following forms:

$$r_s = \frac{A_s - a_g}{A_s + a_g} \tag{3.36}$$

$$r_e = \frac{A_e - a_g}{A_e + a_g} \tag{3.37}$$

3.2.2 Continuum model

The two mass model discussed in the previous section is a simple representation of the vocal folds but gives a good understanding of the mechanics of the vocal folds oscillation. Lumped element models with more masses (i.e. [Story and Titze, 1995]) have been proposed in order to simulate the vocal folds movement in a more realistic way. However, these models had a limitation in the degrees of freedom and could not suggest a detailed movement of the vocal folds [Ikeda et al., 2001]. According to Alipour et al. [2000] the multi mass lumped element models cannot represent the geometry and the viscoelastic behaviour of the vocal folds sufficiently. Therefore, continuum models were suggested and developed, e.g. [Alipour et al., 2000]. The continuum model includes detailed geometrical data and a multilayer vocal folds structure. In Titze and Talkin [1979] a geometrical model of the vocal folds which illustrates the layers of the vocal folds structure is suggested (Fig. 3.10).

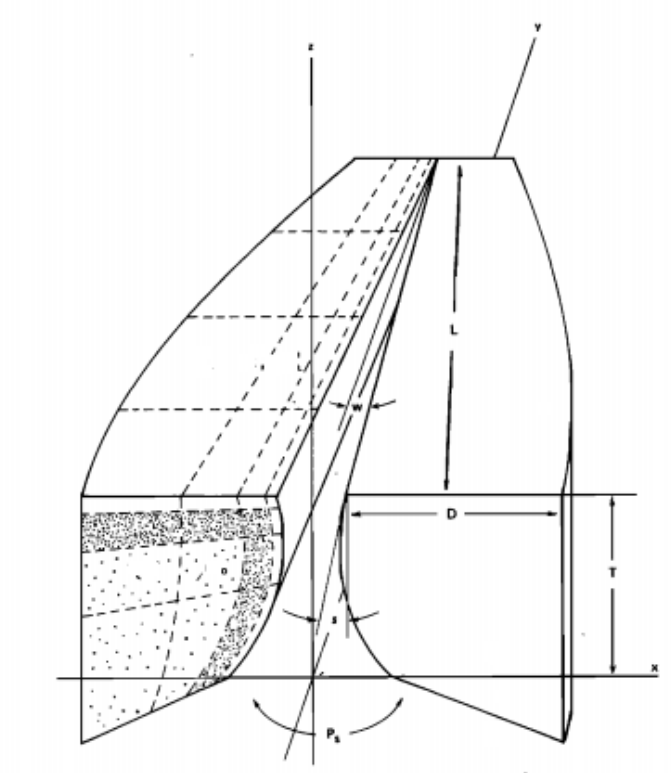


Figure 3.10: Illustration of the vocal folds using the continuum model [Titze and Talkin, 1979].

In Fig. 3.10 the dotted lines are used to separate the tissue compartments. The areas without dots represent the mucosal tissue, the densely dotted area represents the properties of the ligaments and the sparsely dotted area represents the muscular tissue.L is the glottal length, D the depth of the vocal folds and T is the thickness of the vocal folds [Titze and Talkin, 1979].

In order to solve the equations of motion the finite element method is used. Four main assumptions were followed. These are [Titze and Alipour, 2006][Alipour et al., 2000]:

- during vocal folds vibration small deformations occur. Linear elasticity relations are considered.
- single plane vibration is studied.
- the multilayer vocal folds structure is transversely istropic.

A force which is exerted on an area of the body is called *stress*. All the forces that are applied on the body are represented by the *Cauchy's stress tensor* σ. For small displacements, a constitutive equation for six strain and six stress components can be defined as:

$$\begin{pmatrix} \sigma_x \\ \sigma_y \\ \sigma_z \\ \sigma_{xy} \\ \sigma_{yz} \\ \sigma_{zx} \end{pmatrix} = [S] \begin{pmatrix} \epsilon_x \\ \epsilon_y \\ \epsilon_z \\ \epsilon_{xy} \\ \epsilon_{yz} \\ \epsilon_{zx} \end{pmatrix} \tag{3.38}$$

where S is the stiffness matrix. It is a $6x6$ symmetric matrix, which, in the general case, has twenty one independent components [Titze and Alipour, 2006][Alipour et al., 2000]. Though, for transverse isotropic tissue, it has five independent components including the Young's modulus E and the Poisson's ratio ν in he transverse plain, and the Youngs modulus E' , the shear modulus μ', and the Poissons ratio ν' along the longitudinal axis. The shear modulus in the transverse plane can be expressed by the following formula:

$$\mu = \frac{E}{2(1+\nu)} \tag{3.39}$$

According to Granados [2014], if the viscoelasticity is taken into account, the strain given by the constitutive Eq. 3.38 can be written as follows:

$$\sigma(t) = S\epsilon(t) + n\dot{\epsilon}(t) \tag{3.40}$$

where n stands for viscosity and $\dot{\epsilon}$ is the gradient of the strain. Eq. 3.40 and further mathematical manipulation gives the discretized equation of motion:

$$[M]\ddot{x} + [D]\dot{x} + [K]x = [F] \tag{3.41}$$

where $[M]$ is the mass matrix, $[K]$ is the damping matrix and $[K]$ is the stiffness matrix. x defines the nodal displacement and $[F]$ is the forcing vector.

The analytic solution for the continuum model are simple only for one-dimensional deformations and includes the solution of differential equations under certain boundary conditions, as described by Titze and Alipour [2006]. For the three dimensional models, the finite element discretization is implemented.

3.3 Summary

In this chapter, the acoustic characteristics of a tube were discussed by defining the input impedance and the resonances in the vocal tract. The vocal tract was presented as a sequence of coupled tubes which differ in cross sectional area. The frequency domain and the time domain model were discussed. In the time domain model, the vocal tract consists of 44 segments and the length of each segment is $0.3968\,cm$. The latter model was selected to be used for the coupling, because it is easier to couple with the vocal folds model. A two mass model was suggested for the vocal folds and this will be used for the simulations. A short description of the continuum model was also presented. The model includes many unknown parameters that should be defined based on assumptions which could lead to uncertainties. Therefore, including the time restriction, the model was not further analysed.

CHAPTER 4

Geometry of the vocal tract

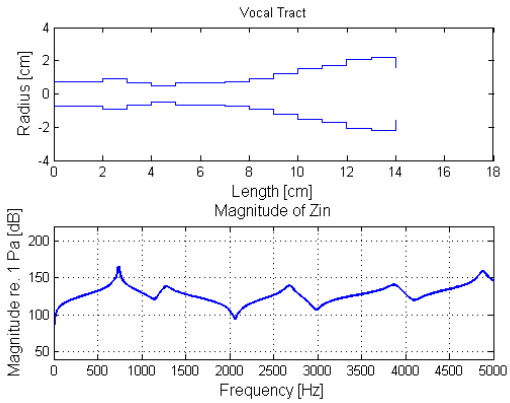
The vocal tract can be approximated as an acoustical tube constituting of multiple coupled segments that differ in cross sectional area. Depending on which vowel is phonated, the vocal tract has a different shape. However, the shape of the vocal tract can differ even if the same vowel is phonated. This suggests that different vocal tract shapes can produce the same acoustic resonance structure [Story, 1995]. Detailed geometrical data are necessary for modelling purposes. More realistic vocal tract shapes can produce a proper distribution of the acoustic pressure across the vocal tract and lead to more realistic formant transitions that approach human speech. For this purpose, many studies used *Magnetic Resonance Imaging or MRI* to obtain detailed geometrical data. MRI is a non invasive and non harmful technique that offers image acquisition [Clement et al., 2007]. Though, there are some disadvantages including that the subject should be in a supine position, the teeth do not appear and the acquisition times are slow. All these can alter the vowel production.

In this chapter, MRI data from Clement et al. [2007] will be used in order to obtain vocal tract shapes and resonances for three spoken vowels. Further, geometrical data obtained from Pedersen [2013] while phonating the four vocal modes will be used to obtain the vocal tract shapes and the resonances for three sung vowels. All the modes where phonated in the same pitch, $261\ Hz$. Moreover, recorded data will be used to investigate the harmonics and the formants of the modes.

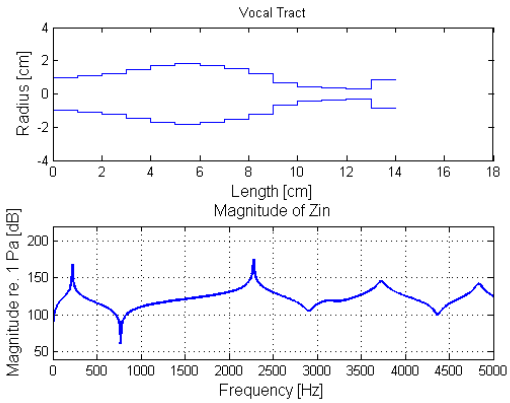
4.1 Simulations using spoken vowels

The vocal tract shape for three different spoken vowels will be discussed in this section. The geometrical data were obtained from Clement et al. [2007] after conducting MRI scans on a male test subject. The vowels which were phonated were $/i/$, $/a/$ and $/u/$.

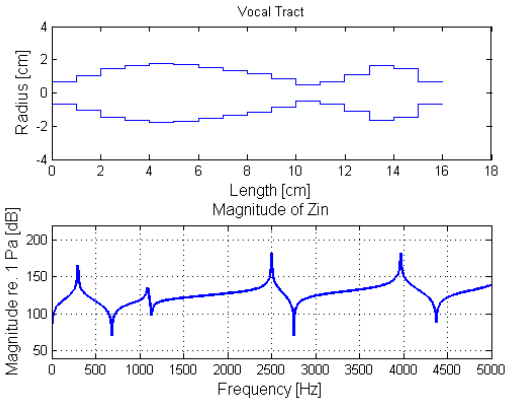
The aim of the this study was to define whether MRI scans can offer a realistic illustration of the human vocal tract. In the current project, the data obtained from the aforementioned study will be used in order to observe the vocal tract shapes and the respective vocal tract resonances. The cross sectional areas were calculated using intervals of $1\,cm$. The vocal tract shape and the input acoustic impedance for each of the three vowels can be seen in Fig. 4.1.



(a) vowel /a/.



(b) Vowel /i/.



(c) Vowel /u/.

Figure 4.1: Vocal tract shape and input impedance for three spoken vowels.

According to Fig. 4.1a, the first four resonances for the */a/* vowel are at $732\,Hz$, $1278\,Hz$, $2692\,Hz$ and $3882\,Hz$ respectively. For the */i/* vowel (Fig. 4.1b), the vocal tract presents the first four resonances at $224\,Hz$, $2278\,Hz$, $3728\,Hz$ and $4834\,Hz$ respectively. For the */u/* vowel (Fig. 4.1c), the vocal tract presents the first four resonances at $296, Hz$, $1090\,Hz$, $2500\,Hz$ and $3968\,Hz$ respectively. In Peterson and Barney [1951] the average of the fundamental frequencies and the formants of 76 speakers is presented. The results from the simulations for the spoken vowels */a/*, */i/* and */u/* show the same trend as found in Peterson and Barney [1951] .

In order to study how the vocal tract resonances behave, the */u/* vowel configuration was selected and compared with the uniform tube configuration. In Fig. 4.2, the vocal tract shapes and the respective resonances of the uniform tube and the */u/* vowel are presented.

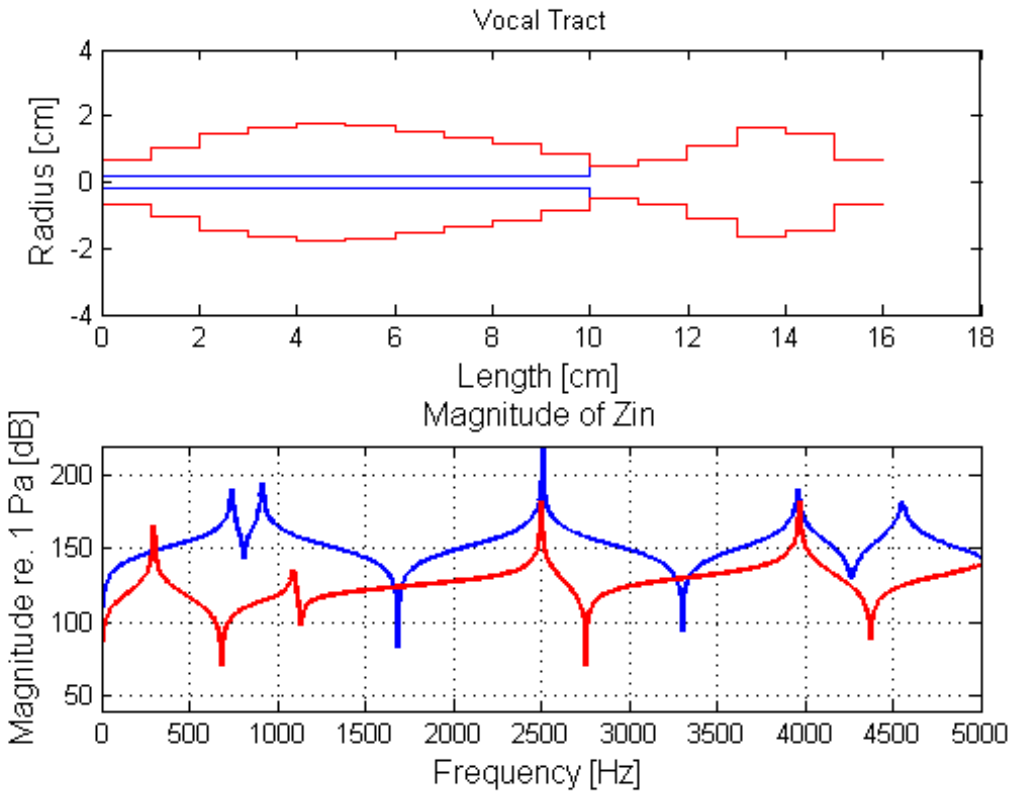


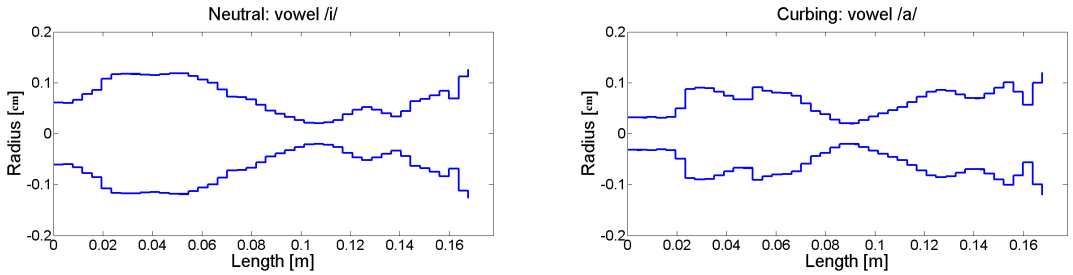
Figure 4.2: Comparison of the vocal tract resonances of the uniform and */u/* vowel.

In Section 3.1.1, it was discussed how a change in the vocal tract shape affects the resonances of the system. A widening of the back part of the */u/* vowel configuration will suggest a shift in the frequencies, meaning a decrease in the first and the third resonances, and a rise in the second and the fourth resonances. This trend can be observed in Fig. 4.1c. The cross sectional area of the uniform tube defines the shift in the resonances. The simulation was repeated for different cross sectional areas of the uniform tube. In all cases the first and the second formant were shifted downwards and upwards respectively. However, the configuration in Fig. 4.2, gave the expected result for all the formants.

4.2 Simulations using the vocal modes

The role of the vocal tract is to filter the signal produced from the vocal folds. The degree of interaction between the vocal tract and the vocal folds is of great significance, because it gives the modes their characteristics. However, the vocal tract shape is important in order to phonate the desirable vowel.

The vocal tract data used in this section were extracted form MRI scans conducted by Pedersen [2013]. During the recordings, a trained singer phonated the four modes. The pitch was the same for the four modes at $261\,Hz$. The vocal tract shapes and the input acoustic impedance were calculated for each mode and for a certain vowel. The results are presented in Fig. 4.3.



(a) Neutral (b) Curbing

(c) Overdrive (d) Edge

Figure 4.3: Vocal tract shape for a) neutral, b) curbing, c) overdrive and d) edge.

The vowels which were phonated during the MRI were $/i/$ for neutral, $/o/$ for curbing and $/e/$ for both overdrive and edge. It is assumed that the length of the vocal tract is $17,46\,cm$. Thus, the acquired MRI data were processed and the vocal tract of each vowel consists of 44 segments. Each segment has a length of $0,3968\,cm$.

In Fig. 4.3, the different vocal tract shapes are attributed not only to the different mode but also, to the different vowel. Though, every mode has a different epilarynx tube shape. Furthermore, it can be observed that neutral and curbing show a narrowing at approximately $9\,cm$ and $11\,cm$ respectively. In addition, edge shows a narrowed tube nearly at the same vocal tract length as neutral and edge. Curbing has the most exquisite narrowing

of all. According to Pedersen [2013], the narrowing occurs at the back of the tongue. In overdrive there is no narrowing. It can also be suggested, that a different phonated vowel could show a less narrowed vocal tract for the neutral mode. The respective vocal tract resonances of each of the aforementioned vocal tract configurations can be seen in Fig. 4.4.

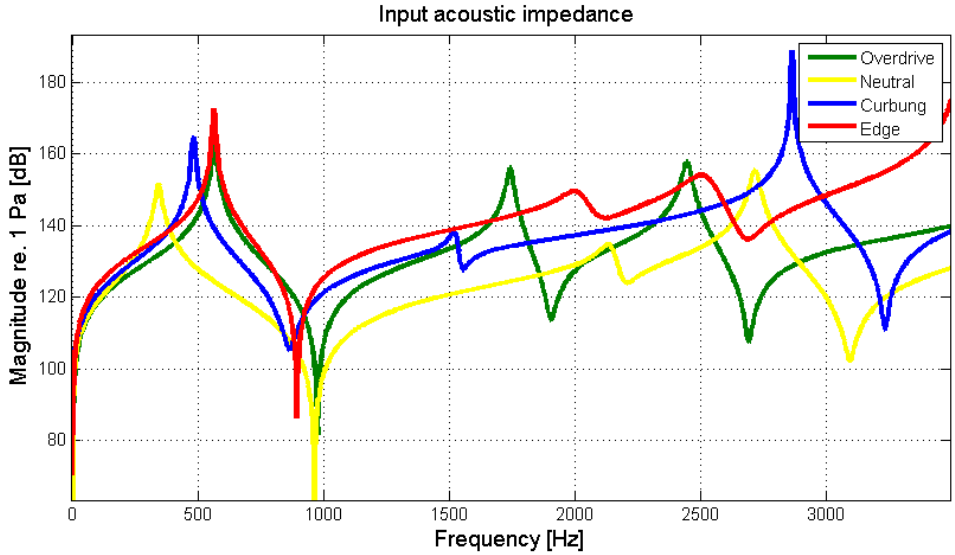


Figure 4.4: Vocal tract resonances of the four vocal modes.

For the overdrive and the edge mode, the same vowel was phonated. This would suggest that the resonant frequencies should have the same values. However, this is not the case for two different modes. Concretely, in Fig. 4.4, it can be observed that for overdrive and edge the first resonance has nearly the same value. Even though the phonated vowel was the same for both overdrive and edge, the different vocal tract configuration shifts the higher formants of edge to higher values. The same trend was observed in [Selamtzis, 2011]. In the latter study, the same vowel was phonated while performing the four modes. The results showed different resonant frequencies for each one of the vocal modes. This verifies that the vocal tract shape is altered while the singer phonates the two different modes.

The calculated resonance frequencies for each vocal tract configuration, according to Fig. 4.4, are presented in Table 4.1.

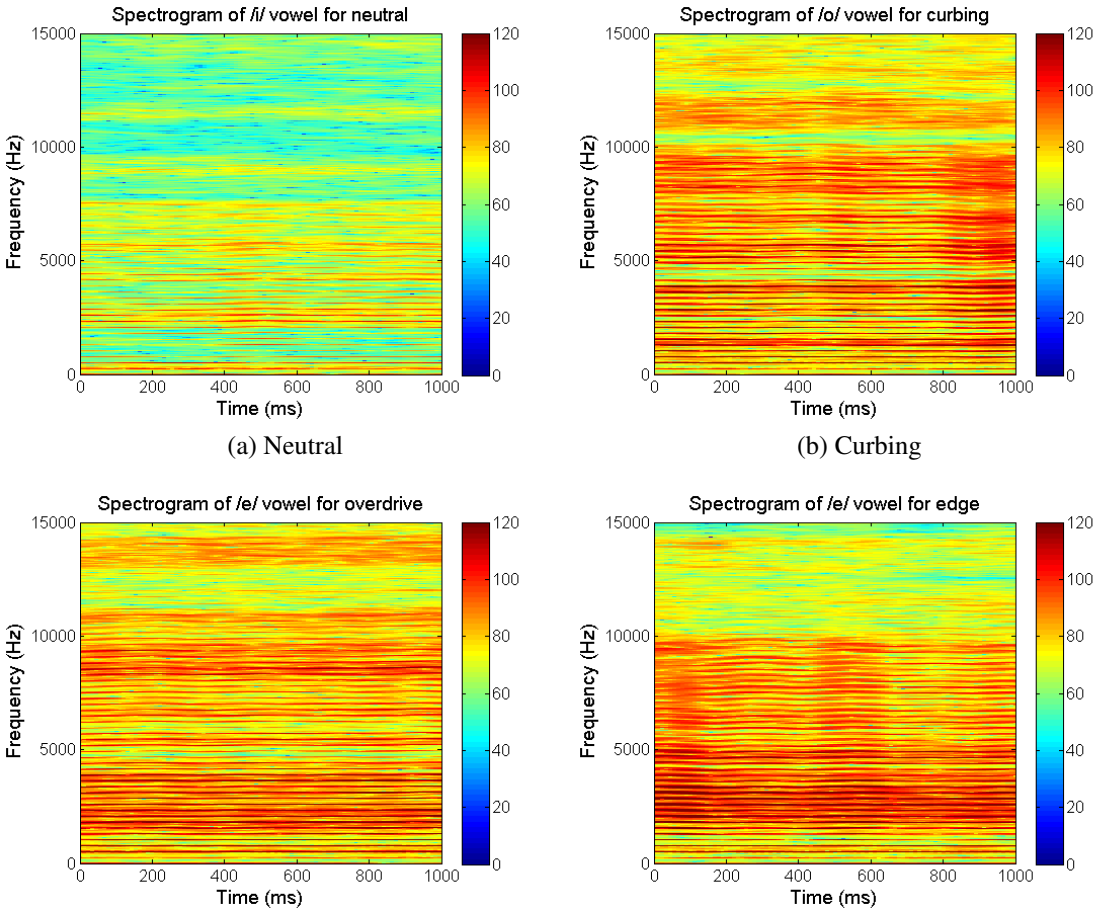
Table 4.1: Resonances of the vocal tract for each vocal mode.

Name of mode	Vowel	R_1 [Hz]	R_2 [Hz]	R_3 [Hz]
Neutral	$/i/$	342	2129	2717
Curbing	$/o/$	482	1522	2863
Overdrive	$/e/$	564	1741	2445
Edge	$/e/$	562	2002	2510

According to Table 4.1 the first formant of the modes shows an increase moving from the non metallic mode to the full metallic modes.

4.3 Analysing the harmonics

An important aspect, which should be taken into account for the modelling of the four modes, is the harmonics of the glottal flow which are produced during the phonation of each of the modes. Therefore, the recorded data, will be processed in order to observe the differences in the harmonics of the modes. The data that were used for the analysis include $1ms$ recordings for every mode. The spectrogram of each mode can be seen in Fig. 4.5.



(a) Neutral (b) Curbing

(c) Overdrive (d) Edge

Figure 4.5: Spectrograms of the four modes a) neutral, b) curbing, c) overdrive and d) edge.

In Fig. 4.5, the red stripes are equally spaced and they represent the harmonics which are produced by the glottal pulse. The scale of the colorbar shows the sound pressure of the harmonics relative to $20\mu Pa$.

According to Fig. 4.5, neutral shows the less strong harmonics compared to the other three modes and the sound pressure of the signal is low. These characteristics verify the low sound level of the mode and the low interaction between the vocal tract and the vocal

folds. Curbing, overdrive and edge show increased sound pressure at harmonics up to approximately $9000\,Hz$. However, overdrive presents strong harmonics for frequencies between approx. $12500 - 15000\,Hz$. The harmonics produced in curbing show a more uniform sound pressure distribution along the spectrum. The narrowing of the epilarynx tube while phonating the three aforementioned modes, suggests a stronger coupling between the vocal folds and the vocal tract. The stronger coupling is accompanied with an increase in the metallic sound of the mode (see Section 2.2).

For a better overview of the formants of the modes, the first four formants were calculated using [Praat, 2010]. The calculated formants are presented in Table 4.2.

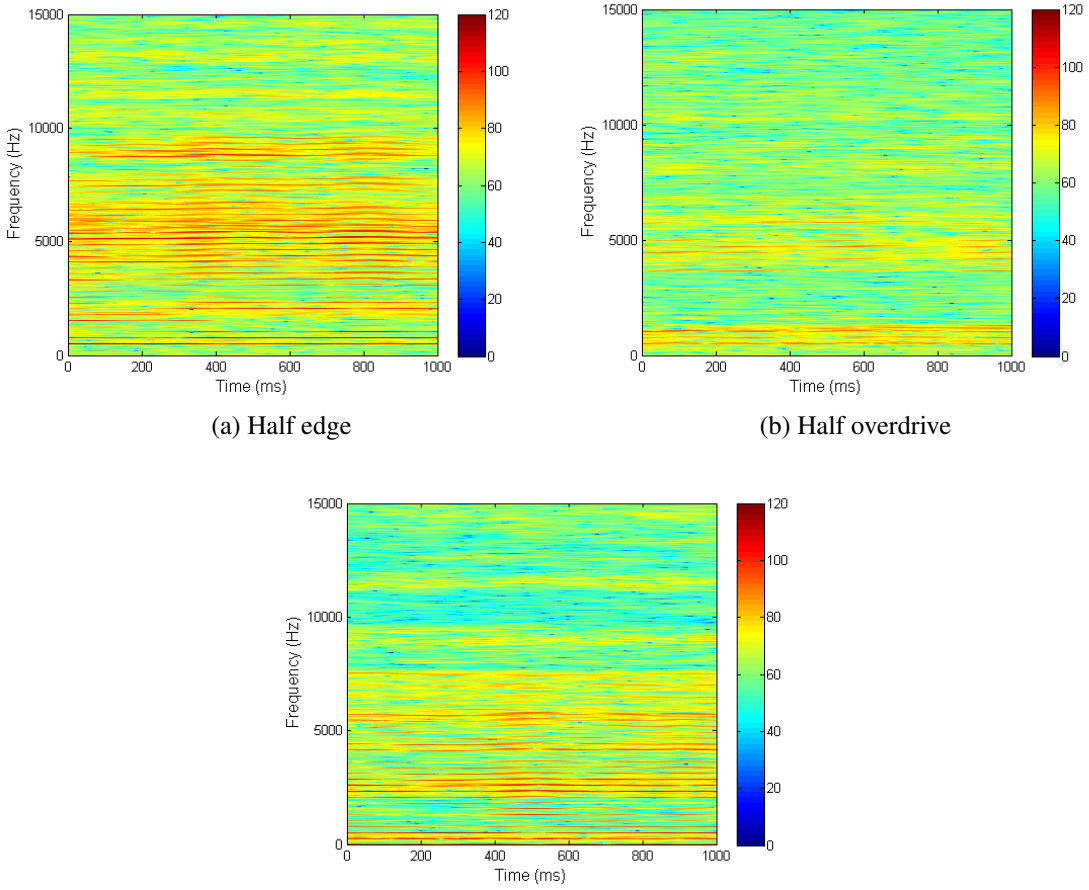
Table 4.2: First four formants of the four vocal modes.

Name of mode	Vowel	F_1 [Hz]	F_2 [Hz]	F_3 [Hz]	F_4 [Hz]
Neutral	$/i/$	286.73	2329.64	2839.48	4236.90
Curbing	$/o/$	713.35	1356.64	2824.91	3841.47
Overdrive	$/e/$	592.15	1871.03	2390.96	3637.73
Edge	$/e/$	630.88	2066.71	2781.26	3353.03

The formants for neutral, overdrive and edge show the same trend as the calculated resonances of the vocal tracts given in Table 4.1. Though, according to Table 4.2, curbing shows increased formant values which can be considered to be in the metallic region. According to Peterson and Barney [1951], the first formant of vowel $/a/$ (like in hot) in average is approximately $730\,Hz$ while the first formant for $/uh/$ (like in but) is approximately $530\,Hz$. Since the MRI scans and the recordings were not obtained at the same measurement, a slight difference in the phonation of the vowel creates this deviation in the value of the first formant of curbing, see Table 4.2 and 4.1.

4.4 Investigation of other possible modes

Recently, curbing is still under research. Half edge, half overdrive, medium edge and medium overdrive are introduced in order to make the characteristics of the modes more clear. Recordings of the modes were processed in order to indicate the differences in their spectra. Half edge and half overdrive spectrograms are illustrated in Fig. 4.6, in comparison with the spectrogram of neutral. All the recordings have the same pitch, $261\,Hz\,(C4)$.



(a) Half edge

(b) Half overdrive

(c) Neutral

Figure 4.6: Spectrograms of a) half edge, b) half overdrive and c) neutral.

According to Fig. 4.6, half edge shows strong harmonics in the frequency range between approx. $4000 - 9000\,Hz$. Half overdrive presents strong harmonics only in the very low frequencies and up to approx. $1000\,Hz$. Let us assume that in half overdrive the epilarynx tube is widened, as in overdrive, see Fig. 4.3. A more widened epilarynx tube in half overdrive, compared to half edge, would suggest less interaction between the vocal tract and the vocal folds, thus leading to less strong harmonics. Compared to neutral, half edge shows stronger harmonics in the whole frequency range. Half overdrive, seems to have less strong harmonics in the low frequency range (up to $5000\,Hz$) but above $7500\,Hz$ it shows more strong harmonics compared to neutral.

Taking into account Fig. 4.6, there is evidence that half overdrive and half edge lie between neutral and the full metallic modes, overdrive and edge. The strength of the harmonics is higher compared to neutral and lower compared to edge or overdrive, see Fig. 4.3.

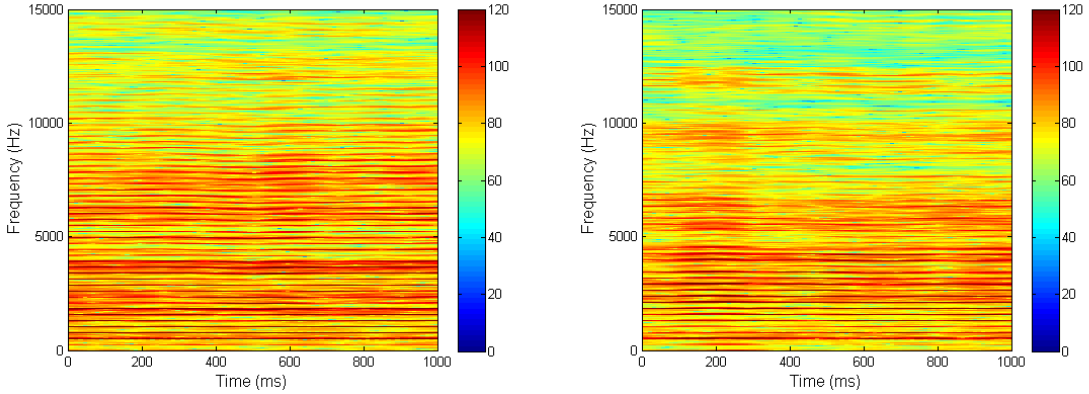
The fundamental frequency and the first four formants for half edge and half overdrive were calculated using Praat [2010]. The results are presented in Table 4.3. Neutral is presented in comparison.

Table 4.3: Formants of half edge, half overdrive and neutral.

Name of mode	F_0 [Hz]	F_1 [Hz]	F_2 [Hz]	F_3 [Hz]	F_4 [Hz]
Half edge	259.49	620.01	1911.61	3563.96	4076.18
Half overdrive	261.08	537.90	1104.13	3237.93	3789.79
Neutral	260.03	286.73	2329.64	2839.48	4236.90

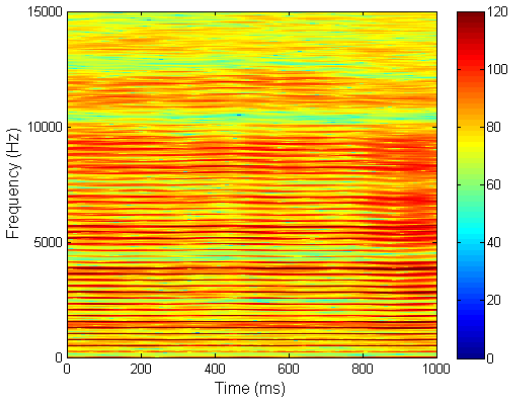
According to Table 4.3 neutral presents the lowest first formant, $286.73\,Hz$. This frequency is close to the pitch, thus it could explain an interaction with the fundamental and stronger harmonics in the low frequency range. For half overdrive and half edge the first formant has higher values compared to neutral, $537.90\,Hz$ and $620.01\,Hz$ respectively. As it was observed in Fig. 4.4, the *metallic* of the sound increases as the first formant increases.

Medium overdrive and medium edge are two modes which are considered to be in the curbing area. The spectrograms of medium overdrive and medium edge are illustrated in Fig. 4.7. Curbing is presented in comparison.



(a) Medium overdrive

(b) Medium edge



(c) Curbing

Figure 4.7: Spectrograms of a) medium overdrive, b) medium edge and c) curbing.

In Fig. 4.7 the three modes were phonated at the same pitch, approx. $261\,Hz$. It can be observed that the strength of the harmonics in medium overdrive and medium edge is increased compared to half overdrive and half edge. This would suggest a more *metallic* sound. However, the medium modes present less strong harmonics in high frequecies ($10000 - 15000\,Hz$) compared to the full metallic modes, overdrive and edge, see Fig. 4.5. Medium overdrive and medium edge show the same trend as curbing. Though, curbing shows a more uniform distribution in the whole frequency range, as far as the strength of harmonics is concerned.

The fundamental frequency and the first four formants for medium overdrive and medium edge were calculated using [Praat, 2010]. The results are presented in Table 4.4. Curbing results are presented in comparison.

Table 4.4: Fundamental frequency and first formants of medium overdrive, medium edge and curbing.

Name of mode	F_0 [Hz]	F_1 [Hz]	F_2 [Hz]	F_3 [Hz]	F_4 [Hz]
Medium overdrive	261.98	540.71	1819.12	2648.11	3595.21
Medium Edge	263.00	542.61	2029.38	2885.58	4066.97
Curbing	257.15	713.35	1356.64	2824.91	3841.47

According to Table 4.3 medium overdrive and medium edge have approximately the same value for the first formant at $540.71\,Hz$ and $542.61\,Hz$ respectively. The phonated vowel defines the shape of the vocal tract and consequently the formants. The phonated vowel for curbing was different from the two other modes and this justifies the high value for curbing's first formant. However, the spectrograms in Fig. 4.7 show the same trend for the three modes. It can be suggested that medium overdrive, medium edge and curbing belong to the same category.

Summarizing the results

In this chapter data from spoken and sung vowels were processed. It can be seen that the formants for the same spoken and sung vowel differ e.g. for $/i/$ vowel. In each case, this is attributed to the different vocal tract configurations. For instance, Fig. 4.1b and Fig. 4.3a show different vocal tract shapes even though the same vowel was used.

It was observed that the strength of the harmonics among the four differs. Neutral shows the less strong harmonics. This could be justified by the widened epilarynx tube that makes the interaction between the vocal tract and the vocal folds weak. Thus, the low sound level of the mode can be explained. The two full metallic modes, overdrive and edge, show strong harmonics in the low and medium frequency range (approx. up to $9000\,Hz$). Curbing presents strong harmonics in the whole frequency range. Edge has increased sound pressure in up to $5000\,Hz$. The presence of increased sound pressure in these two modes seems to be due to the narrowed epilarynx tube which implies a stronger coupling.

In the last part of this chapter, the investigation of the behaviour of four new suggested modes was presented. It seems that half edge and half overdrive are two transient modes between neutral and the two full metallic modes. Medium edge and medium overdrive

present similar harmonic distribution as edge and overdrive respectively but, it could be said, that they show less strong harmonics.

It should be noted that the results presented in this section were derived from a limited number of test subjects. Therefore, more recordings from different subjects would suggest a more representative conclusion about the differences among the modes.

CHAPTER 5

Further modelling and coupling

Unlike an ideal lossless tube the yielding walls of the vocal tract impose losses on the system when they vibrate [Story, 1995]. In this chapter, the modelling of the vocal tract will be continued including the losses due to the vibration of the yielding walls. Furthermore, the coupling of the vocal folds model with the model of the vocal tract will be discussed. The models that are used are the ones described in Chapter 3.

5.1 Losses in the vocal tract

In Chapter 3, the vocal tract was modelled as a sequence of uniform tubes with different cross sectional areas. The only losses that were taken into account was the radiation impedance at the end of the tube (i.e. the mouth). However, in order to create a more stable model, especially for the case of the coupled model, the losses in the vocal tract due to the vibration of the yielding walls are important. Let us assume a two tube junction. According to Story [1995], the vibration of the walls is represented as a lumped element circuit, as seen in Fig. 5.1. The parameters that are used for the illustration in Fig. 5.1 are the following:

- L_w, represents the mass-like properties of the flesh.
- C_w, represents the pressure storage capabilities.
- R_w, represents the viscosity of the tissue.

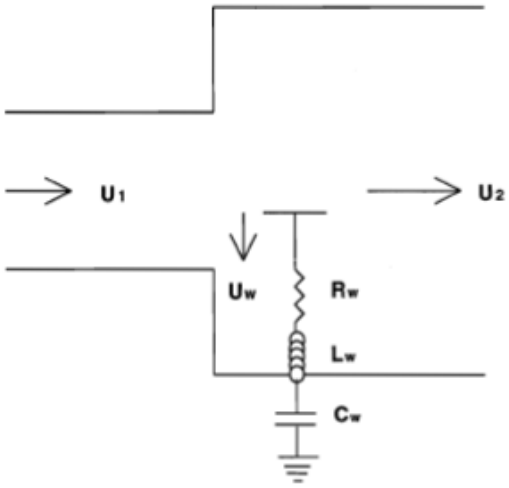


Figure 5.1: Lumped element circuit representing losses due to the yielding walls [Story, 1995].

In Fig. 5.1, U_w represents the flow into the lumped element circuit. The effects of the yielding walls and the derivation of the new scattering equations are mentioned in Story [1995]. The vocal tract resonances were calculated using the new scattering equations and are illustrated in Fig. 5.2.

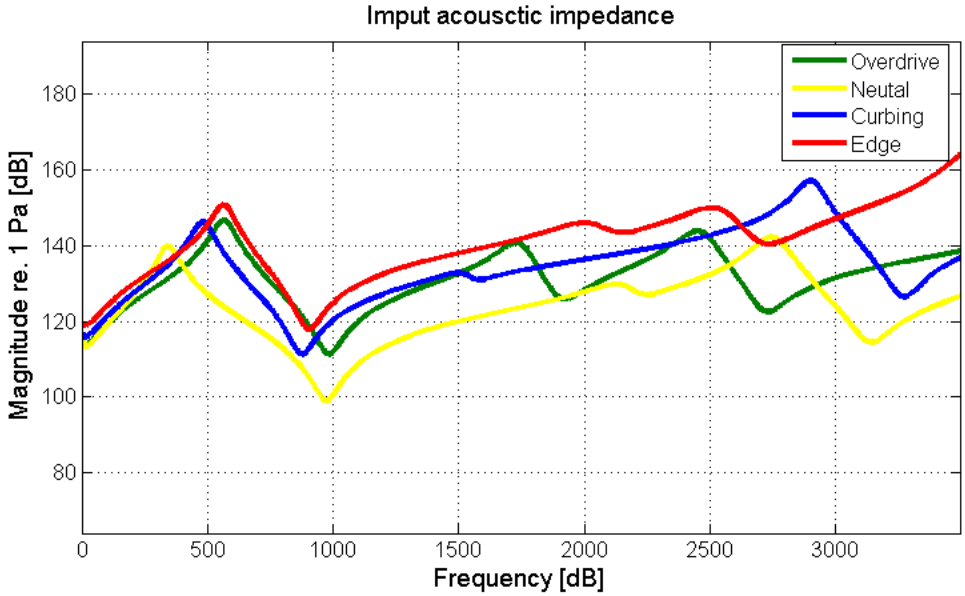


Figure 5.2: Input acoustic impedance of the four modes with losses.

Fig. 5.2 can be compared with Fig. 4.4. In Fig. 4.4, it can be observed that in a lossless vocal tract the resonances show infinite amplitude. However, Fig. 5.2 show smoother input acoustic impedance curves along the frequency range. Thus, for further simulations and modelling a lossy vocal tract will be taken into account.

5.2 Results using the vocal folds model

In Chapter 3 a two mass vocal fold model was suggested. In this section, the parameters of the model were adjusted in order to produce a desirable pitch by using a tension parameter q. an increase in the tension parameter means an increase in fundamental frequency. For the first simulation the tension parameter was $q = 1$ with a fundamental frequency at $140\,Hz$. The sampling frequency was $44100\,Hz$ and the pressure was kept constant at $800\,Pa$. The displacement of the two mass were $x_{1,rest} = 0.001\,m$ and $x_{2,rest} = 0.001\,m$ respectively. The results of the vocal folds model using the parameters in Table 3.1 are shown in Fig. 5.3.

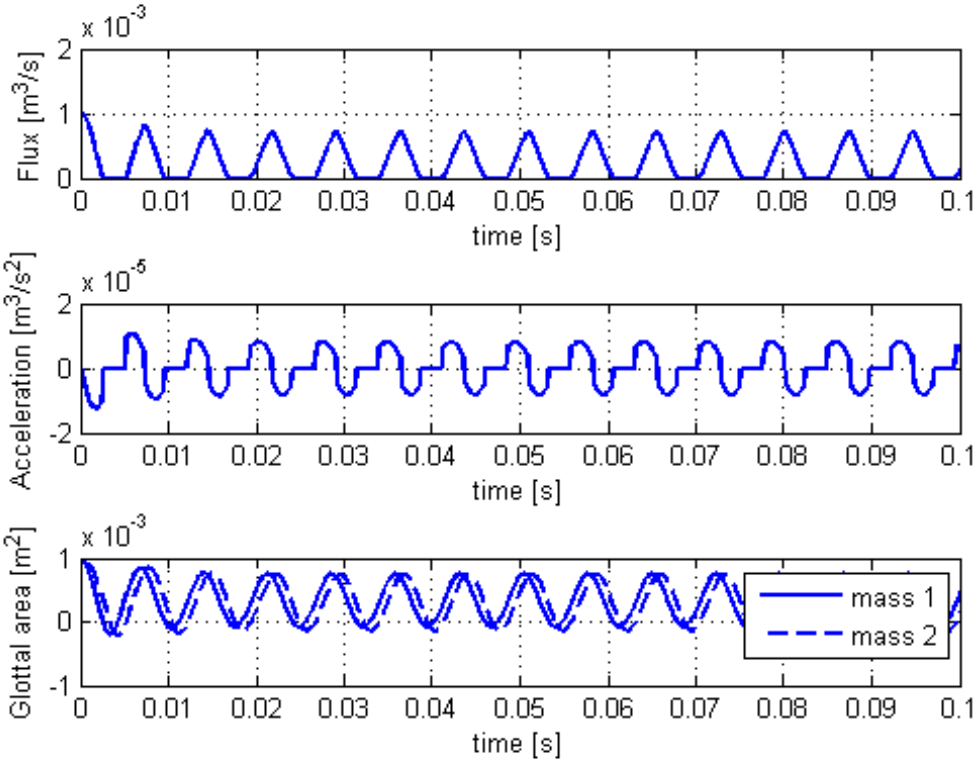


Figure 5.3: Flux, acceleration and time varying glottal area of the two masses of the vocal folds.

In Fig. 5.3 a self oscillated system of the vocal folds can be seen. This means that the oscillation of the vocal folds is self sustained without the vocal tract loading. When the glottal area takes negative values it means that the vocal folds are in the collision area. The lower and upper masses m_1 and m_2 show a difference in phase. The upper mass is driven by the lower mass and this results in a phase delay of the upper mass. According to Titze [1988] the wave propagation in the mucosa is from bottom to top. Reversed

situation would not lead to oscillation. Therefore, the lower mass should always drive the upper mass.

5.3 Coupling the vocal tract

In order to couple the vocal tract with the vocal folds model a subglottal and supraglottal tract were created. The subglottal tract is a configuration consisting of multiple coupled uniform tubes with the same cross sectional area. The length of the human subglottal tract is approximately $14.3\,cm$. Thus, the number of the coupled tubes was set to 36 in order to obtain a length of $0.3968\,cm$ for each of them. The supraglottal tract, as mentioned in the previous chapters, consists of 44 segments. Each of them has a length of $0.3968\,cm$ in order to obtain a total length of $17.46\,cm$ which is approximately the length of the human supraglottal tract.

As far as the vocal folds model is concerned, it should be noted that the latter model needs further improvement in order to represent the real vibration of the vocal folds. Further, during the coupling the model is very sensitive since a small shift in the initial displacements of the two masses can lead to a damped system. The pitch of the recordings which were investigated in Chapter 4 was $261\,Hz$. The pitch with the current vocal folds model was adjusted to approximately $264\,Hz$ by selecting a tension parameter $q = 1.6$. Adding only the subglottal tract the flow in the first tube of the vocal tract, as seen from the glottis, can be seen in Fig. 5.4.

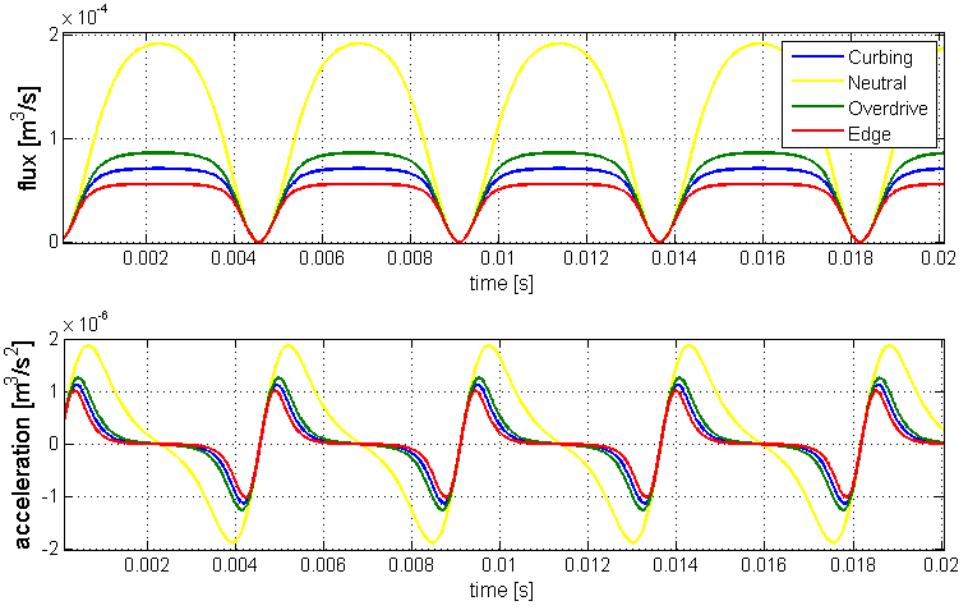


Figure 5.4: Flux and acceleration at the first tube of the supraglottal tract.

In Fig. 5.4 it can bee seen that neutral presents the maximum flow with overdrive, curbing and edge following respectively. This trend is attributed to the fact that neutral presents a widened epilarynx tube compared to the other modes, thus letting more air to pass through the vocal tract.

When the subglottal tract is included in the system the flow plays an important role. Concretely, first, the flow will start increasing and gradually will be stabilized. When the reflected pressures from both the subglottal and supraglottal tract show great differences, the Bernoulli effect is present. This should alter the movement of the vocal folds. In Fig. 5.5 the flow and the volume acceleration during the coupling are presented.

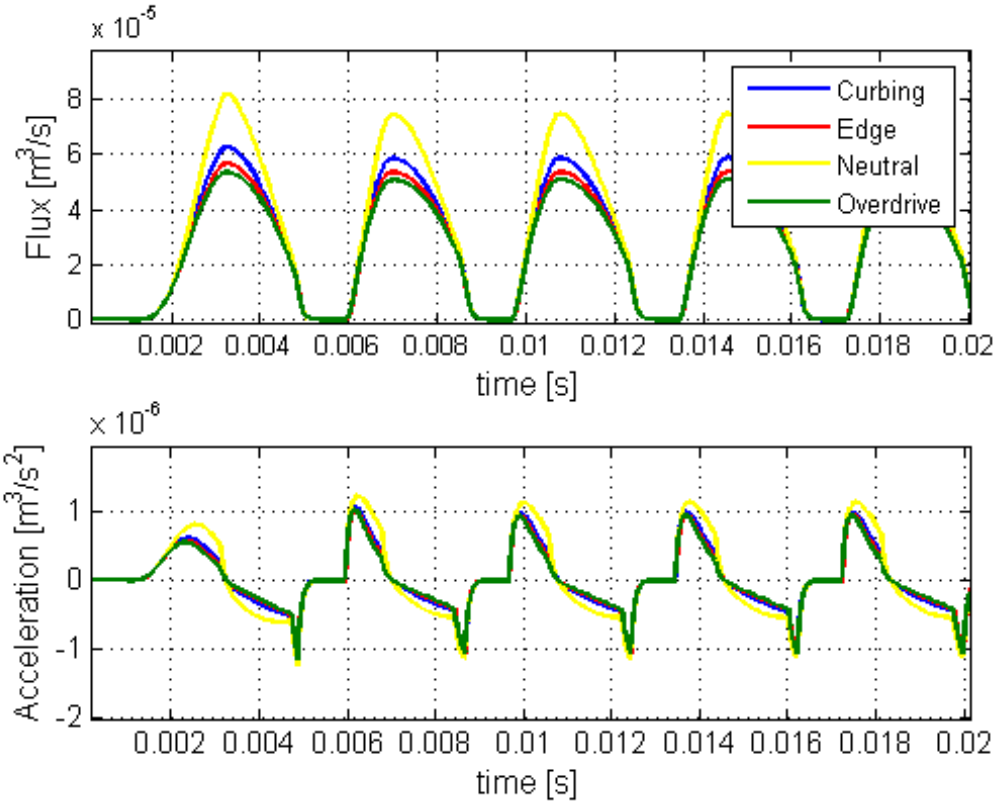


Figure 5.5: Flux and acceleration for a pitch at $264\,Hz$.

In Fig. 5.5 it can be seen that in the beginning of the oscillation the flow increases and after some seconds is stabilized. However, the shape of the flow pulses is kept the same for the three modes. This should not be the case for the coupled system [Pedersen, 2013]. Furthermore, the flow pulses in the coupled case should be amplified compared to Fig. 5.4. This could be explained by the fact that the initial displacement of the two masses in the coupled case was much smaller compared to Fig. 5.4. Furthermore, according to Pedersen [2013] the displacement of the masses in combination with the tension parameter can simulate pressed or normal phonation. In order to obtain the desirable pitch in

the coupled model, both the tension parameter and the displacement of the masses were changed. Thus, Fig.5.4 and Fig. 5.5 present different phonation cases and this can explain the difference in the behaviour of the two systems.

In order to see the behaviour of the system for a higher pitch, the tension parameter was set to $q = 2$. The obtained pitch was $387\ Hz$. The displacement of the masses was the same as in the case for the $264\ Hz$ simulation. The results are illustrated in Fig.5.6

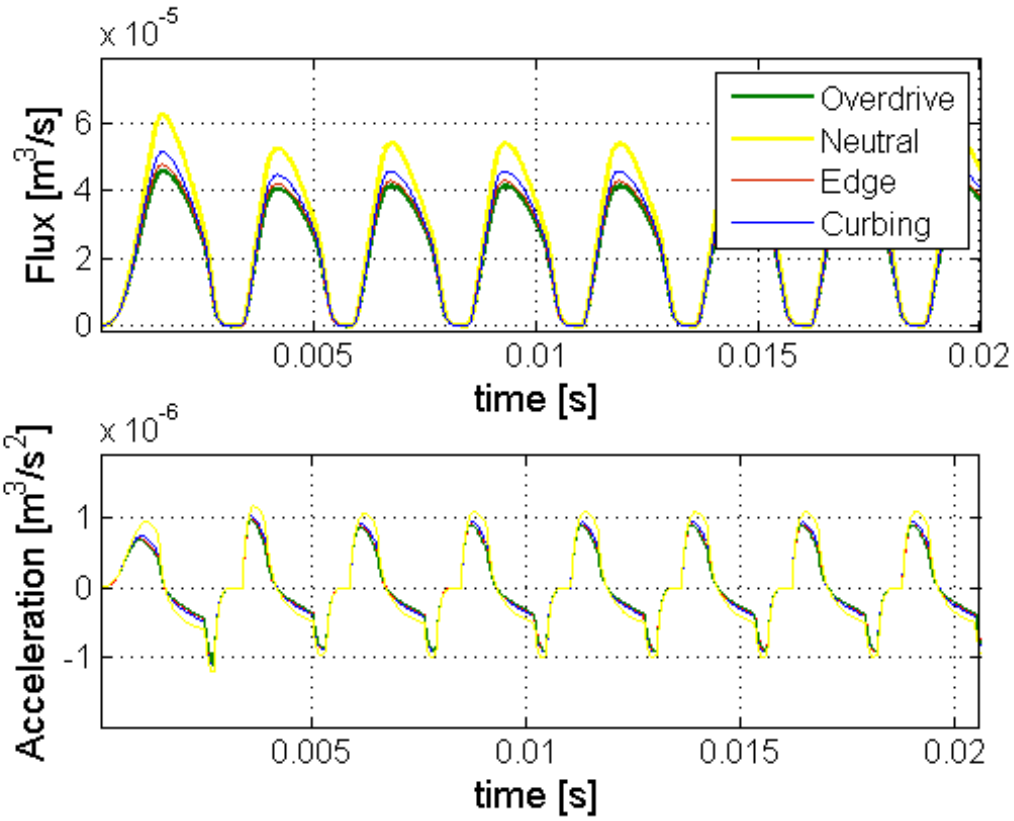


Figure 5.6: Flux and acceleration for a pitch at $387\ Hz$.

In Fig.5.6, it can be observed that a rising in the pitch leads to less flow for all the modes. The increase in the tension parameter influences the mechanical movement of the vocal folds. The forces that are included in the equations of motion 3.23 decrease. Additionally, the solution of the equations of motion includes these forces, as explained in Appendix B, in order to calculate the displacement of the lower and upper mass. Indicatively, the displacement of the two masses, as well as the minimum glottal area for the $387\ Hz$ and the $264\ Hz$ pitch are presented in Fig.5.7 for the case of neutral.

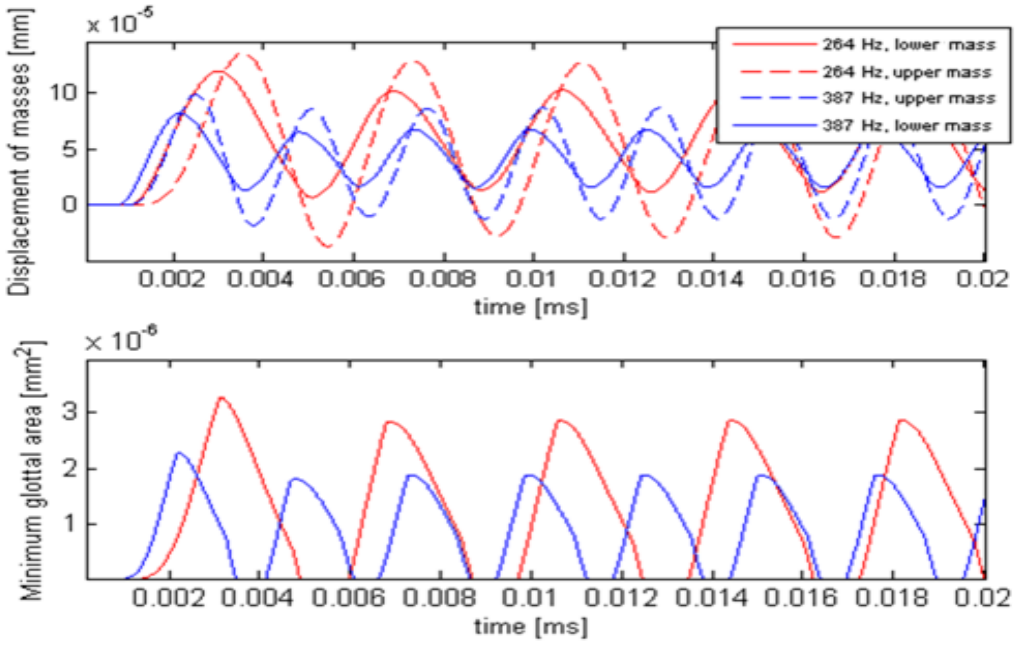


Figure 5.7: Displacement of the upper and lower mass for two different pitch values.

In Fig. 5.7 it can be seen that for both cases the lower mass drives the upper mass. This satisfies the condition for the wave propagation in the mucosa, according to Titze [1988] as mentioned in Section 5.2. The effect of the pitch shift can also be observed. In low pitch the oscillation has a high amplitude and the minimum glottal area is higher compared to the high pitch. The difference in the tension parameters makes the system oscillate in different amplitude. The higher the stiffness the more difficult for the masses to oscillate.

The reflected pressures were calculated for the four modes and the results are illustrated in Fig. 5.8.

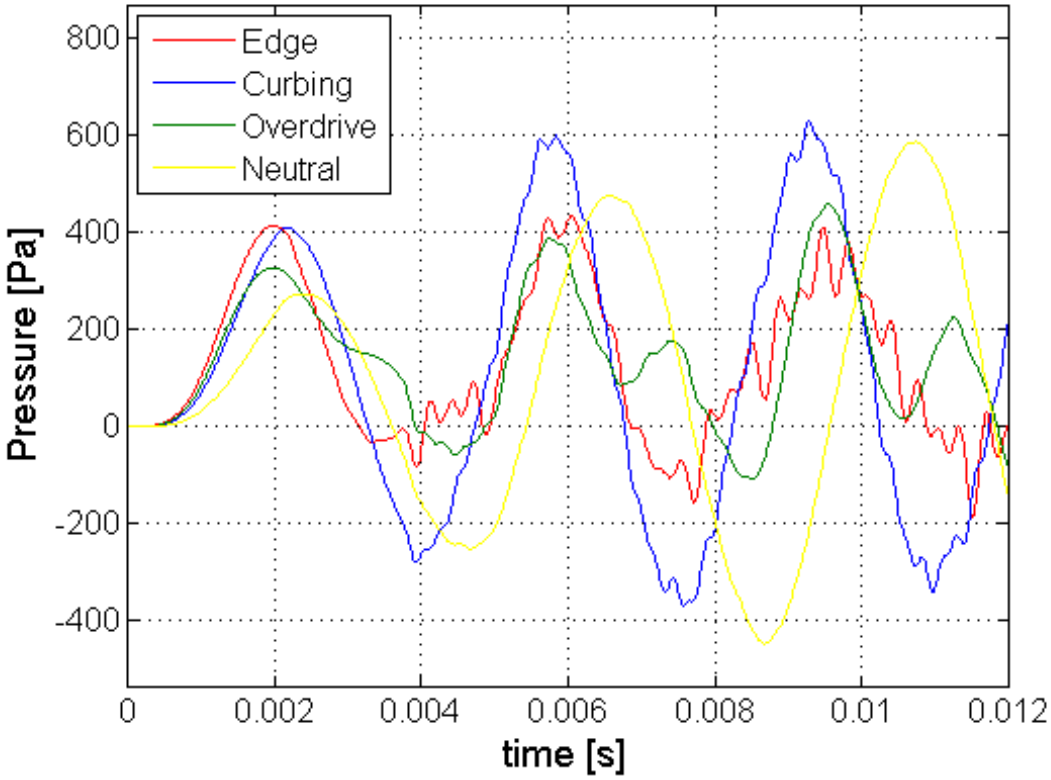


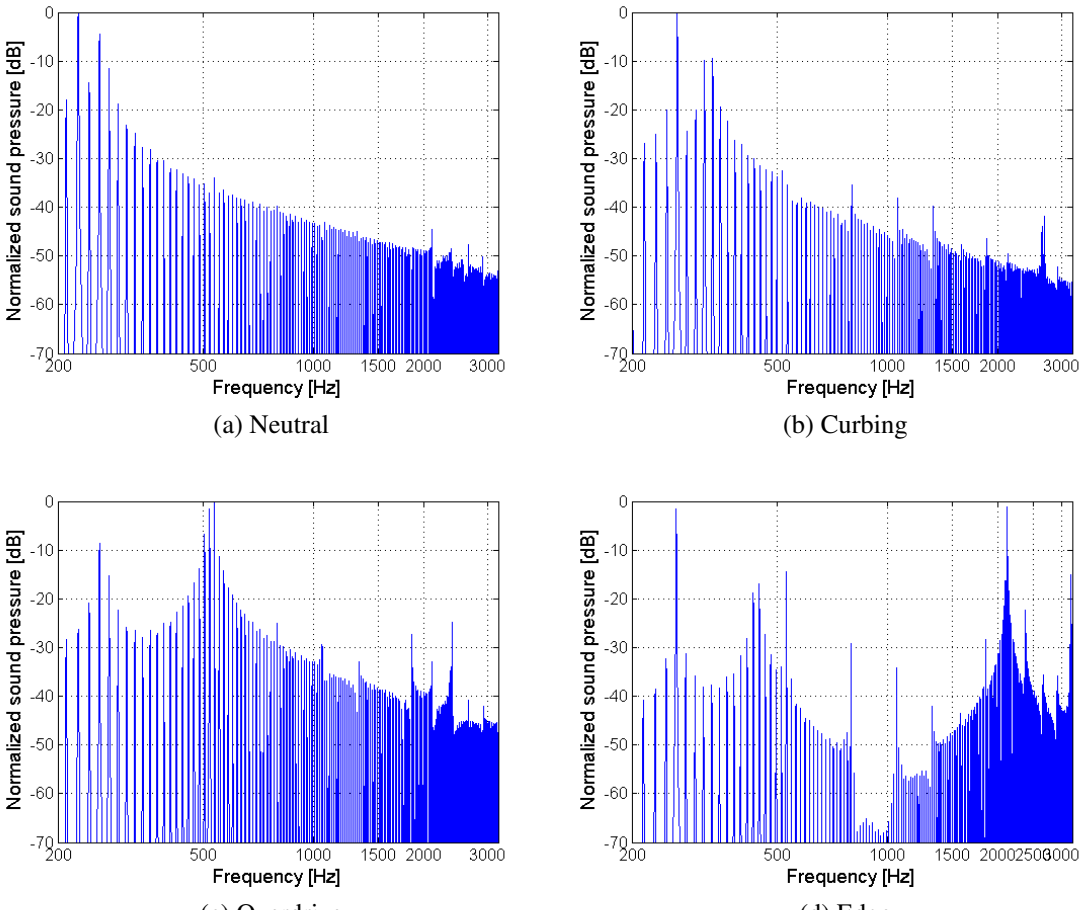
Figure 5.8: Reflected pressures during the coupling.

According to Fig. 5.8 the reflected pressures in the vocal tract, start from a low amplitude value and gradually they increase. However, neutral tends to increase in amplitude more than the other three modes which is not the expected trend. Furthermore, in Fig. 5.8, a shift in neutral's cycle can be seen. The same trend is observed in Pedersen [2013], where neutral shows the same inertive behaviour but the cycle is shifted. This is related to the fact that the fundamental frequency have similar values with the first formant, where the resistive load increases [Pedersen, 2013].

It is worth noticing that the pressure at the beginning of the glottal cycle is positive. This indicates the inertive load of the vocal folds. According to Titze [1988], the air in the vocal tract acts like a mass of air accelerated and decelerated. During the opening phase of the vocal folds the flow increases (see Fig. 5.5) and the air in the vocal tract increases with glottal air. For the acceleration a positive pressure is needed. At the closing phase the reduced suppraglotal pressure lowers the average glottal pressure and at some point the average intraglottal pressure becomes negative [Titze, 1988]. The fluctuations that are observed in the pressure curves are attributed to the higher harmonics that resonate in the

vocal tract [Pedersen, 2013].

The next step of the analysis was the investigation of the harmonics and define their spectrum. A fast Fourier transform was applied on the volume velocity of the obtained signals using a hamming window. The sound pressure of each harmonic was divided by the value of the strongest harmonic in order to get the normalized amplitude. The spectrum of the four modes are depicted in Fig. 5.9. Note that the spectra are zoomed in order to compare the strength among the harmonics of each mode. The obtained spectra are listed in Appendix C in comparison to the spectra calculated using the recorded data used in Section 4.2.



(a) Neutral

(b) Curbing

(c) Overdrive

(d) Edge

Figure 5.9: Spectrum of a) Neutral, b) Curbing, c) Overdrive, d) Edge.

In Fig. 5.9, the strong peaks are the harmonics and are integer muliples of the fundamental frequency. It can be observed that in neutral, the spectrum presents a strong peak before the fundamental frequency, approx. at $228\ Hz$. The presence of the overtone could be either due to the non-linear oscillation of the masses or due to the phase difference of the produced waves.

In neutral's spectrum, the decrease in the sound pressure as the harmonics increase is in accordance with the results obtained in Section 4.3. Neutral show the less strong harmonics in the whole frequency range. The same trend is observed for curbing but this cannot considered a valid result since strong harmonics are expected in low and high frequency range. In the overdrive spectrum, the second harmonic shows an increased sound pressure compared to the fundamental (approx. $8.77\,dB$ difference). In Titze [2008], it is mentioned that the amplitude of the harmonics is influenced by the reactance curve. The first harmonic of overdrive according to Fig.5.9 is approximately at $529.4\,Hz$ and at the positive reactance area. This could increase the amplitude of the harmonic. In edge, the fourth harmonic shows a more increased sound pressure. The fourth harmonic is in the positive reactance area for edge, which could imply an increase in the amplitude of the harmonic.

Overall, the obtained results from the model discussed in this chapter are not in complete accordance with the results obtained from the analysis of the recorded data in Section 4.3. Many of the harmonics are underestimated which would suggest a more accurate investigation of the system, by including more parameters. Important aspects that have to be taken into account are the initial displacement of the masses, the tension parameter, the open quotient and the pressed phonation. In this study, the model was simplified and some of the above mentioned parameters were not examined thoroughly. The symmetrical vocal folds model and the fact that the initial positions of the masses were always kept the same could lead to deviations from the measurements.

CHAPTER 6

Conclusion and further research

A vocal tract and a vocal folds model was implemented in order to examine their interaction. A simplified two mass model was used for the modelling of the vocal folds. The current project was based on two previous studies concerning the singing voice and the investigation of the four vocal modes [Selamtzis, 2011] [Pedersen, 2013].

Firstly, an analysis of the resonances in the vocal tract of the four modes was conducted. For this purpose, already existing MRI data were used. Every mode shows a different vocal tract shape. However, the different vocal tract shapes are also attributed to the different vowels that were phonated during the measurement. The input acoustic impedance was the main metric that was used during the simulations, and the resonances of the vocal tract were calculated. It seems that moving from the *non metallic* modes to the *full metallic* modes, the first resonance of the vocal tact increases.

Furthermore, an approach to find possible differences among the modes was made by analysing recorded data. It was observed that the strength of the harmonics is different among the modes. It can be assumed that, the stronger harmonics a mode has, the *metallic* sound of the mode increases. Investigation of four new suggested modes were also conducted. It appears that between the *non metallic* modes and the *full metallic* modes there are two intermediate modes. These modes could offer a soft transition from one mode to another.

It should be noted that the number of the recordings was limited. Thus, a more valid investigation of the modes would include an experiment where several well trained singers will phonate vowels using the different modes. It would be of great interest if the singers used the same vowel for each mode.

The final step was to model the modes. During the coupling of the two models it is evident that there is an interaction between the two models. Though, the obtained results did not match with the results obtained from the recordings. A more detailed vocal folds model, using more parameters, would probably yield in more representative results. Due to time restrictions and the complexity of the physiology of the vocal system a more detailed approach was not performed.

Overall, it should be mentioned that a very complex system was modelled in a very simple way. More focus should be given on the modelling of the vocal folds. The continuum model that was not used for further processing in this study, can be used for a more realistic vocal folds model. Furthermore, EGG measurements in combination with recordings would give more information about the open quotient of every mode, especially for the lately discovered modes. Last but not least, it would be worth investigating whether the singers use the formant tuning technique when they perform the *full metallic* modes.

Appendices

APPENDIX A

Matlab code used for the WR analog

```
l=0.1746;%length of vocal tract
n_end=43;%number of vocal tract segments

S=S_NT;%cross sectional area

rho=1.2;%air density
c=343;%speed of sound
fs=((n_end*c))/(l);%sampling frequency
Ts=1/fs;%length of the sample

n_end=n_end+1;
for n=1:n_end-1
r(n)=(S(n)-S(n+1))/(S(n)+S(n+1));
end

F=zeros(1,n_end);
B=zeros(1,n_end);
F_2=zeros(1,n_end);
B_1=zeros(1,n_end);
Fsave=zeros(1,n_end);
Bsave=zeros(1,n_end);

a=sqrt(S(end)/pi);
R=128/(9*pi^2);
L=2*fs*8*a/(3*pi*c);
a_2=-R-L+R*L;
```

```
27 a_1=-R+L-R*L;
28 b_2=R+L+R*L;
29 b_1=-R+L+R*L;
30
31 r_glot=1;
32 t_end=32800;
33 for t=1:t_end;
34
35     if t==1
36         u=1;
37     else
38         u=0;
39         F=F_2;
40         B=B_1;
41     end
42
43     for n=1:1:n_end-1
44         F_2(n+1)=(1+r(n))*F_2(n)-r(n)*B_1(n+1);
45         B_1(n)=r(n)*F_2(n)+(1-r(n))*B_1(n+1);
46     end
47
48 F_2(1)=u*rho*c/S(1)+r_glot*B(1);
49 B_1(n_end)=1/b_2*(F_2(n_end)*a_2+F_2(n_end)*a_1+B_1(n_end)*
       b_1);
50
51 Fsave(t,:)=F_2;
52 Bsave(t,:)=B_1;
53 end
```

APPENDIX **B**

Solution of the equations of motion

In order to solve the equations of motion in Section 3.2.1 the forward Euler method was followed according to Agerkvist [2013]. Let us assume that:

$$\dot{x_1} = u_1 \tag{B.1}$$

$$\dot{x_2} = u_2 \tag{B.2}$$

The equations of motion 3.23 and 3.24 can be written as follows:

$$\dot{u_1} = \frac{1}{m_1}\left(F_{e1} - F_{1col} - r_1 u_1 - k x_1 - k_c((x_1 - x_{1,rest}) - (x_2 - x_{2,rest}))\right) \tag{B.3}$$

$$\dot{u_2} = \frac{1}{m_2}\left(F_{e2} - F_{2col} - r_2 u_2 - k x_2 - k_c((x_1 - x_{1,rest}) - (x_2 - x_{2,rest}))\right) \tag{B.4}$$

The general form of the solution is written as:

$$x(n+1) = x(n) + T_s(Fx(n) + G) \tag{B.5}$$

where T_s is the time of the interval, F is a $4x4$ matrix and G is a $4x1$ that presents the non linear part of the F matrix. F and G are matrices defined by Eq. B.3 and B.4. Let us assume that $x_{1,rest}$ and $x_{2,rest}$ are equal to zero. After proper mathematical manipulation the matrices F and G are defined as follows:

$$F = \begin{pmatrix} -r_1/m_1 & 1 & (-k_1 - k_c)/m_1 & k_c/m_1 \\ 1 & -r_2/m_2 & k_c/m_2 & (-k_2 - k_c)/m_2 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix} \tag{B.6}$$

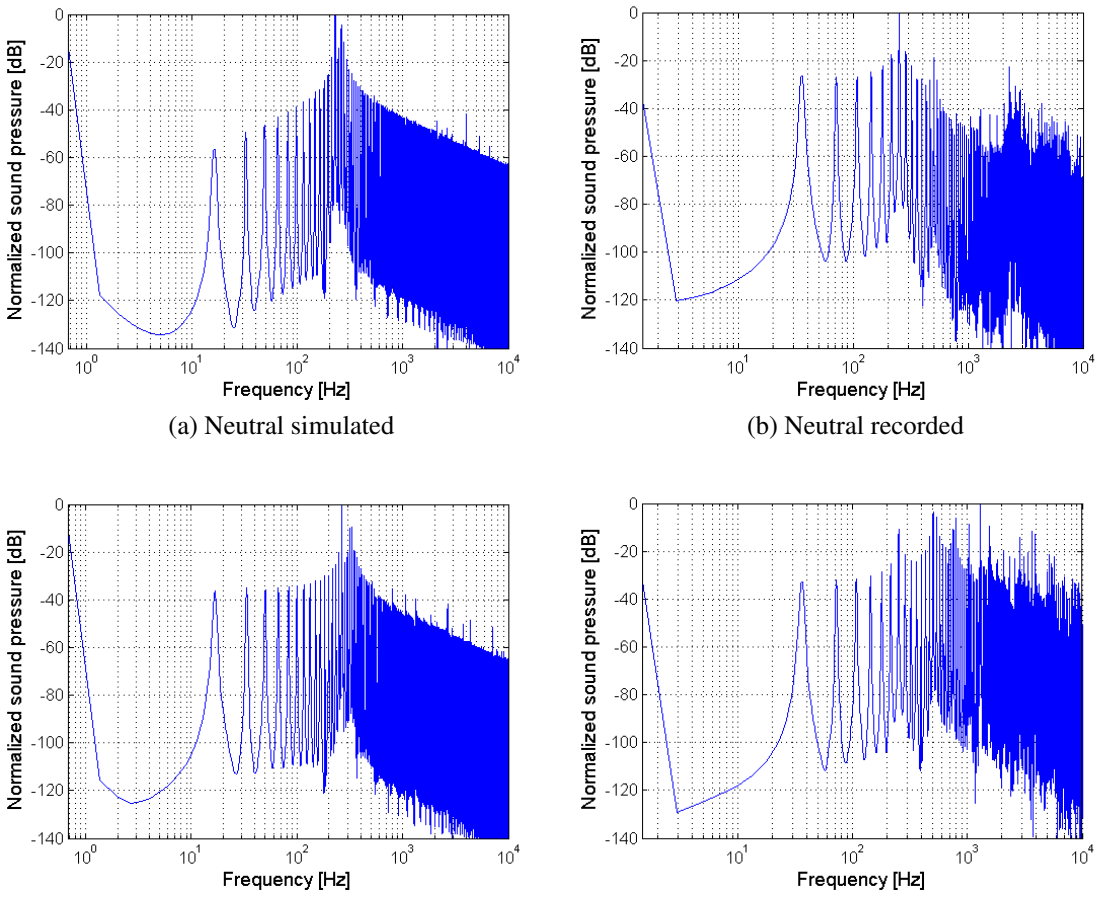
and

$$G = \begin{pmatrix} \frac{F_{e1} - \Theta(-a_1) * c_1 * \frac{a_1}{2l_g}}{m_1} \\ \frac{F_{e2} - \Theta(-a_2) * c_2 * \frac{a_2}{2l_g}}{m_2} \\ 1 \\ 1 \end{pmatrix} \tag{B.7}$$

APPENDIX C

Simulated vocal modes: Spectra

The spectra of the four vocal modes were obtained by applying a fast Fourier transform and a hamming window. The results of the simulations are presented in comparison to the spectra of the recorded signals. They can be seen in Fig. C.2



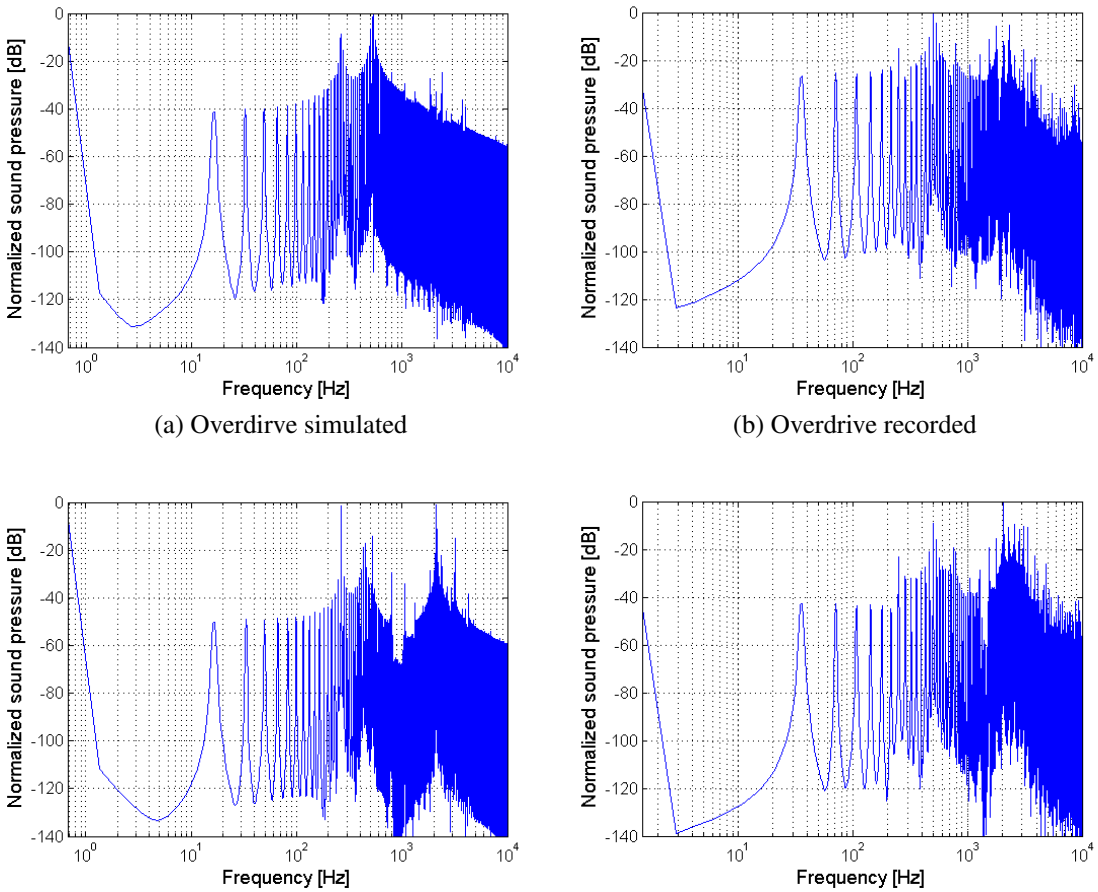
(a) Neutral simulated

(b) Neutral recorded

(c) Curbing simulated

(d) Curbing recorded

Figure C.1: Spectrum of neutral a) simulation, b)recording, and curbing c) simulation, d) recording.



(a) Overdirve simulated

(b) Overdrive recorded

(c) Edge simulated

(d) Edge recorded

Figure C.2: Spectrum of overdrive a) simulation, b)recording, and edge c) simulation, d) recording.

Bibliography

Agerkvist, F. [2013], 'Discrete time modelling of nonlinear loudspeakers'.

Alipour, F., Berry, D. A. and Titze, I. R. [2000], 'A finite-element model of vocal-fold vibration', *Acoustical Society of America* **1**08(2), 3003–3012.

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

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.

Chang, K.-h. [2012], Speech Analysis Methodologies towards Unobtrusive Mental Health Monitoring, PhD thesis, EECS Department, University of California, Berkeley.
URL: *http://www.eecs.berkeley.edu/Pubs/TechRpts/2012/EECS-2012-55.html*

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.

completevocalinstitute.com [2015], 'Complete vocal institute @ONLINE'.
URL: *http://completevocalinstitute.com/research/laryngograph/*

Fakotakis, N. [2005], *Speech Technology*, University of Patras Press.
URL: *http://www.GetBodySmart.com*

Fant, G. [1960], *Acoustic theory of speech production*, Mouton and Co.

Flanagan, J., Allen, J. and Hasegawa-Johnson, M. [2008], *Speech Analysis Synthesis and Perception*.

GetBodySmart.com [2015], 'Getbodysmart.com @ONLINE'.
URL: *http://www.GetBodySmart.com*

Granados, A. [2014], 'Biomechanical models of the vocal folds (i)', *Lecture note for course Advaced acoustics* .

Hanayama, E., Camargo, Z., D.H., T. and Rebelo Pinho, S. [2009], 'Metallic Voice: Physiological and Acoustic Features', *Journal of Voice* **1**(3), 62–70.

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.

Ikeda, T., Matsuzaki, Y. and Aomatsu, T. [2001], 'A Numerical Analysis of Phonation Using a TwoDimensional Flexible Channel Model of the Vocal Folds', *Journal of Biomechanical Engineering* **1**23, 571–579.

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], 'Greens function in a room', *Exercise guide used in the course "Advanced Acoustics" at DTU spring 2012* .

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

NCVS.org [2015], 'The national center for voice and speech @ONLINE'.
URL: *http://www.ncvs.org/*

Pedersen, J. [2013], 'The singing voice: A vibroacoustic problem', *DTU* .

Peterson, G. and Barney, H. [1951], 'Control Methods Used in a Study of the Vowels ', *Acoustical Society of America* **2**4(2), 175–184.

Plack, C. [2005], *The sense of hearing*, Press.

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

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

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

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.

StudyBlue.com [2015], 'studyblue.com @ONLINE'.
URL: *https://www.studyblue.com/notes/note/n/larynx/deck/12423026*

Sundberg, J. [1994], 'Perceptual Aspects of Singing', *Journal of Voice* **8**(2), 106–122.

tdmu.edu [2015], 'Respiratory system: anatomy and physiology @ONLINE'.

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. [1976], 'On the mechanics of the folds vibration.', *J. Acoust. Soc. Am.* **6**0(6), 1366–1380.

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. [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 Talkin, D. [1979], 'A theoretical study of the effects of various laryngeal configurations on the acoustics of phonation.', *Acoustical Society of America* **6**6(1), 60–74.

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 [2015], 'Larynx @ONLINE'.
URL: *https://en.wikipedia.org/wiki/Larynx/media/File:Illu_larynx.jpg*

List of Figures

List of Tables