Snoring Sound Production and Modelling

Acoustic Tube Modelling

by

Mert Ergin

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Tuesday September 27, 2018 at 10:30 AM.

Student number:	4627849	
Project duration:	December 1, 2017 – August 3, 2018	
Thesis committee:	Dr. Ir. R. Heusdens,	TU Delft, supervisor
	Dr. Ir. B. Brinker,	Koninklijke Philips N.V.
	Dr. Ir. J.C.A. van der Lubbe,	TU Delft

This thesis is confidential and cannot be made public until September 31, 2021.

TU Delft

Confidentiality Clause

This thesis with the title *Snoring Sound Production and Modelling* is based on internal and confidential data of Philips Research. This work may only be available to the reviewers and authorized members of the board of examiners. An inspection or publication or duplication of this thesis - even in part - requires the expressed authorization of the author or the management of Philips Research. This confidentiality clause is temporally restricted and expires three years after the submission date of the thesis.

Preface

In the beginning of my master programme, Electrical Engineering - Signal & Systems at TU Delft, I understood the depth and sophistication of signal processing. Interested in acoustics and speech production, thanks to my father, I decided to follow courses on audio signal processing. During Digital Audio and Speech Processing course at Technical University of Delft in 2017, I was intrigued with the fact that sound production and perception is deeply intertwined with human anatomy.

The curiosity led me to Dr. Ir. Richard Heusdens who talked about projects at Philips Research, and I contacted him immediately. Thankfully, I got a response in short time from Bert den Brinker and Koray Karakaya from Philips Research in Eindhoven on a project about snoring. As snoring has been a part of my childhood, again thanks to my father, I was interested. Berts endeavour to guide me with his extensive knowledge throughout my work and his never-ending patience towards me can not be appreciated enough here. Additionally, Okke Ouweltjes' assistance in acoustic theory and software aspects of the project helped me a lot on understanding a topic that I do not have any background in. Richard Heusdens' understanding and kind attention in all matters helped conducting my thesis at Philips Research Eindhoven successfully. I will be grateful to my supervisors as their wisdom has inspired me to achieve greatness. Consequently I want to thank Dr.ir. JCA van der Lubbe for being a part of the thesis committee.

Foremost, I want to thank my partner, Asligul Keskin, my family and especially my parents, for their never-ending support throughout this Master thesis.

The thesis at hand is structured into five chapters: The Introduction chapter gives a brief summary of the state of art of snoring sound analysis and the anatomy of the upper airway. Furthermore the thesis objective is presented here. Chapter two explains the theoretical background of speech production and linear prediction analysis and how this problem relates to snoring sound production. The experiments and results section involves the application of the theory and difference of models suggested in the previous section. Results and limitations will be discussed in chapter four. The thesis ends up with a conclusion part and proposals regarding the future work on the topic.

Contents

List of Figures

List of Tables

1 Introduction

It is likely that man has found the snoring of others unpleasant since cave dwelling times, but it was not until the second half of 20th century that effective medical and surgical treatments were developed. Exact figures on the prevalence of snoring are difficult to give, as the experimental results depend highly on the chosen population based on ethnic origin and age. Despite studies show different prevalence, It is generally assumed that around 40% of the adult population snore habitually [20][27][51][50][80][77].

The sound of snoring consists of a series of impulses caused by the rapid obstruction and reopening of the upper airway. This cycle of closure and opening occurs in the region of 50 to 130 times per second during a snore event[47][12]. In most cases, loud snoring does not pose any health threat although it is a frequent cause of marital disharmony. In 10% of habitual snorers, complete collapse of the airway during sleep leads to the Obstructive Sleep Apnea (OSA) [32]. Due to adverse effects on the cardiovascular system and daytime tiredness, this disease has been linked with higher than average rates of death from both accidental and natural causes. In the last twenty years there has been an increasing awareness of the problems of loud snoring and OSA. As a consequence, doctors have been seeing a much greater demand for medical and surgical treatment.

Similarly, speech generation is enabled by opening and closing of the vocal cords in the vocal tract and speech synthesis has always been a fundamental area in speech processing. It refers to the process of producing an acoustic signal by controlling a model with a set of parameters. Similar to speech, snoring sounds are produced by vibrating tissues within the upper airway tract. By exploiting this similarity, we try to explain snoring sound production using synthesis techniques that has been applied to speech.

1.1. Anatomy of Snoring Sound Production

In order to define different types of snoring, we have to be acquainted with anatomy of the upper airway tract. Snoring occurs in sleep, during which the upper airway is in a passive state. As most of the snoring sounds are produced during inhalation [68], the driving pressure is directed towards the interior. These physiological differences should be kept in mind when applying speech techniques to the acoustic analysis of snoring. In this section, we will describe the inhalation process, and mention the tissues and organs that play a role on inhalation.

By moving the thoracic diaphragm muscle, a suction effect is created at the lungs. Air may either move through the mouth or the nose or through both simultaneously. The nose is a rigid walled tube with bony walls and hence is not prone to collapse.

We will focus on the supra-glottal area, as it involves anatomical parts that are larger contributors to snoring. These parts are soft tissues that tends to collapse during sedation and

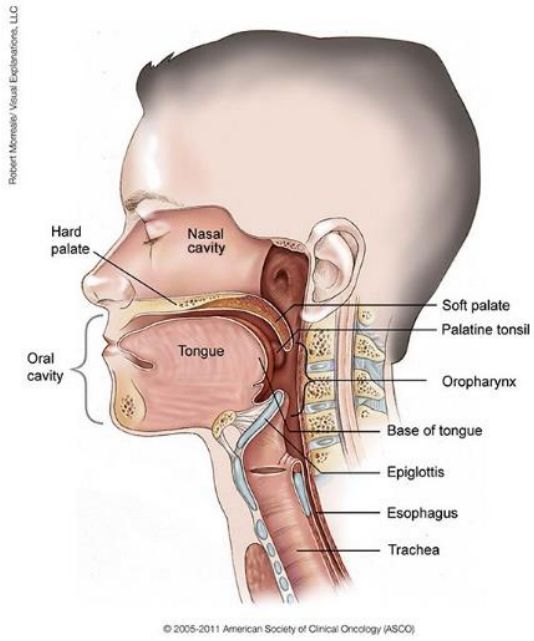


Figure 1.1: Sagittal View of the Upper Airway [65]

sleep. These four structures, described in Kezirian et al.'s study are Velum, Oropharyngeal Walls, Tongue and Epiglottis i.e. VOTE structure.

First, the velopharyngeal area is the area that is located around the velum, which is a soft tissue located at the back roof of the mouth. Attached to the velum, we have uvula and around the velum, we have lateral velopharyngeal walls, which are similar to oropharyngeal walls. Velum is also called the soft palate as it does not constitute a bone. Velum functions as a flap or valve that closes off the nasal airway during swallowing to prevent the regurgitation of food into the nose.

The oropharyngeal lateral walls are located around the palatine tonsils. Lateral collapse of these tissues also induce sleep disordered breathing. While located in the same area as the velum, the impact can be distinguished in both audio signals and endoscopic investigation [53].

Another structure that can cause sleep disordered breathing is the tongue. Similar to speech the tongue can change the cross sectional area of the structure by moving backwards towards the anteroposterior wall, causing obstruction in the airway. This is caused by relaxation of the muscles during both sedation and several sleep phases such as Non-Rapid Eye Movement (NREM)and Rapid Eye Movement (REM). This muscle tone relaxation, supported by the impact of the gravity during supine sleeping position. As a result, the tongue moves to the posterior region, causing the cross sectional area to decrease, and in turn enables snoring.

The upper airway tract also includes the epiglottis which is an elastic cartilage that is attached to the entrance of the larynx. The epiglottis separates the trachea and the esophagus and controls air and liquid intake during breathing and swallowing respectively. Epiglottis is one of the possible contributors to snoring. It is found that epiglottis is an under-recognized factor in sleep disordered breathing [53]. The epiglottis can collapse from two reasons: decreased rigidity of the structure and displacement to posterior wall.

The sub-glottal area consists of larynx, trachea and lungs. Larynx and trachea contain a strong framework of cartilage and are not prone to collapse under normal circumstances. Collapsing of the lungs (Pneumothorax) is a different problem that does not lead to snoring.

These four structures can have different levels of obstruction. Kezirian et al. [53] suggested a scheme for diagnosing sleep disordered breathing in Table 1.1 on the next page in 2011. At the posterior side of the airway tract, we have posterior pharyngeal wall. One of the collapsing mechanism is collapsing of the aforementioned regions to posterior wall. This

mechanism is called antero-posterior collapse. Another mechanism is called lateral, which refers to collapsing of the structures to the center of the airway. The last mechanism is called concentric, which is combination of antero-posterior and lateral collapse.

Table 1.1: VOTE classification scheme suggested by Kezirian [53], Open boxes reflect potential degree of obstruction, shaded boxes reflect the fact that a specific structure-configuration cannot be observed

Structure	Degree of Obstruction	Configuration		
		Antero-Posterior	Lateral	Concentric
Velum				
Oropharynx Lateral Walls				
Tongue Base				
Epiglottis				

Snoring is caused by the coupled oscillation of the walls of the airway with the airflow travelling through it. From the point of view of mechanics, these are problems related to the stability of flows over flexible structures. In the study of fluid-structure interactions, linear analyses play a vital role, although quantitative results call for the inclusion of non-linear effects. Linear analysis investigates the behaviour of very small disturbances to an otherwise steady state. As a result fluid dynamics is an alternative approach to modelling of the upper airway and snoring problems, and is briefly mentioned in the next section.

1.2. Literature Review

This section will go over previous literature on snoring sound production and summarize how these research contribute to current state of the art.

Huang et al.[47] investigated soft palate snoring in experimental settings in 1995. They used an experimental model to simulate inspiratory flow through the pharynx. As a result, they have found that below a critical flow speed, the soft palate remains still. Once the speed is exceeded, the soft palate flaps back and forth, hitting the sides of the airway tract. Photographic evidence reveals that the flap of the soft palate starts as small-amplitude oscillations which increase in size over several cycles until the tube wall is hit. This conforms with the concept of flutter between the tongue and pharyngeal wall.

First, we go over approaches that involve sound features. It is mandatory to mention that most of these studies and many others are focused on OSA classification and segmentation of the sounds that are produced by humans during sleeping. Earliest research on snoring sounds were conducted by Miyazaki et al. [59] in 1998. Recorded snoring sounds have been analyzed and compared with intraluminal pressure measurement and dynamic image recordings. The results show that sound recordings are good for screening or supportive method to diagnose the site and grade of upper airway obstruction during sleep. In 2006, Duckitt et al. demonstrated that hidden Markov models and acoustic parameterizations, which are effective in speech recognition systems, are able to identify snores in audio recordings with high accuracy [32].

In 2007 Cavusoglu et al. suggested an algorithm that classifies sleep sound segments as snores and non-snores according to their subband energy distributions [23]. As they observed similarities between snoring energy distributions they used Principal Component Analysis (PCA) to classify the sleeping sounds in two dimensional space and compared the results with Ear Nose Throat (ENT) experts classifications. They have found that below 500 Hz frequency region is sufficient to classify snoring and non-snoring segments. In 2011, Azarbazin

and Moussavi [8] used an unsupervised learning method, fuzzy clustering, which adjusts the parameters for the algorithm adaptively and used a V-Box approach. In 2012, Ben-Israel et al.[13] classified snore and non-snore using an acoustic-based system and proved inter-snore properties can differentiate adult habitual snorers according to Apnea-Hypoapnea Index (AHI). Doukas et al. [30] proposed a signal processing based approach and estimated Sound to Noise Ratio (SNR) and noise level using Likelihood Ratio Testing (LRT) and classified sound frames as snore/non-snore. Similarly Dafna et al. [33] classified snoring using variety of spectral and temporal measures and used a feature selection algorithm on the Receiver operating characteristic (ROC) curve to select most definitive features. Karunajeewa [52], Alshaer [5], Flanagan [37] also suggested similar approach for detecting snoring sounds. Alshaer[6] also focused on segmenting inspiration and expiration sounds. Most of these methods were applied to patients in a silent room with recording equipment. Additionally, the main focus of these studies are on the spectral features that has been driven directly from the snoring sounds. Moreover, anatomical structure of the airway has not been incorporated in the solution of the problem and the detecting where the problem has not been the objective of these studies.

More recently, the team in Passau University worked on data evaluated by ENT experts based on classification suggested by Kezirian et al. [53]. The team worked mainly on classification of different snore types obtained from different patients. Later, in the INTERSPEECH 2017 Computational Paralinguistics Challenge Snoring subchallenge [75], The Munich-Passau Snore Sound Corpus (MPSSC) is introduced for classification of snore sounds by their excitation location within the upper airways. Based on video and audio recordings from three different ENT centers taken during drug induced sleep endoscopy (DISE) examinations, snore events have been identified and classified into four classes based on a simplified version of the VOTE classification. The resulting dataset, MPSSC, contains 828 snore events from 219 subjects. While this dataset would be useful for our project, it was inaccessible throughout the project duration. However, it shall be pointed that the recordings of this corpus are taken in well controlled environment, with equipment, and observation and annotation of medical experts which we lacked in our study.

It is necessary to mention that endoscopy is used in combination with sedation (Drug Induced Sleep Endoscopy (DISE)) in most of the aforementioned studies. Agrawal et al.[2] mentions that drugs can have an effect on tissue stiffness. Therefore, it is questionable to use snoring sounds collected during DISE for studying typical everyday snoring conditions. In our study, patients will not be sedated and will be emitting natural sounds during their sleep.

Another approach to snoring problem is investigating the airway structure in mechanical domain. Aittokolio et al. [4] built a model to predict nasal flow profile that considers three critical components which control the patency of the airway; stiffness of tissues, pump drive and muscle support. Balint et al. [11] built a computational model to simulate the motion of an arbitrarily deforming cantilevered flexible plate embedded in viscous plane-channel flow to identify the mechanisms when breathing impairment occurs through snoring and sleep apnea. Chouly et al.[26] proposed an in vitro experimental model in order to simulate behaviour of the airway structure and its interaction with flow. Lastly, Babbs [10] simulated movement of the soft palate as a pendulum moving with varying pressure caused by the Venturi effect in nasal and oral part of the airway. While these methods do stem from physical phenomena and has the objective to find the flow patterns within the mouth, they do not incorporate the sound emitted from the snorer. They mainly focus on the anatomical structure and the pressure pattern within the airway and model the flow patter for the vibration during the inhalation. We, on the other hand, analyze real life snoring sounds, incorporating the upper airway anatomical structure in the solution.

The study that is more in line with our theoretical model is conducted by Ng et al. [62] [63] and [61] between 2006 and 2010, who worked on speech like analysis of snoring sounds. Additionally, cross sectional areas of upper airway have been investigated [64]. He used inverse filtering methods to find the source flow and later, used cross-sectional areas to synthesize

snoring sounds. Later, he conducted a listening test and measured if narrowed airways can create snoring sounds which are perceptually different. More recently, Saha et al.[73] worked on developing a subject-specific acoustic model of the upper airway to investigate the impacts of upper airway anatomy.

To summarize, many approaches have been used to alleviate snoring problems, assist medical personnel for surgical diagnosis, help annotation of snoring sounds, create a reliable physical model and many other problems related to snoring. Our objective will be explained in the following section.

1.3. Goals of the Thesis

Many different approaches, ranging from sound feature based, to liquid flow modelling has been used for snoring problems. Our main approach will stem from speech synthesis theory. Therefore, the main objective of the present thesis is to answer the following questions.

1. Can acoustic tube modelling be used to analyze the upper airway tract and used to locate the obstruction?
2. Are linear prediction coefficients reliable for reconstructing the upper airway dimensions for snoring sounds?
3. Can excitation flow obtained from inverse filter be a significant indicator of which area is obstructed in the upper airway?
4. Is it possible to identify the severity of obstruction in non-OSA patients using Source Filter Model (SFM)?
5. Is it possible to identify the location of obstruction in non-OSA patients using Source Filter Model (SFM)?

1.4. Outline

In this chapter we have explained the anatomy of snoring sound production and introduced previous research on snoring sounds. Then, in Chapter 2 we will explain the theoretical background of speech production starting from the wave equations and build our models based on transmission line theory and inverse filtering. The following Chapter 3, will be devoted to experiments and results for our models. Next, in Chapter 4 the results from the experiments will be discussed and the limitations of the study will be described. Consequently, Chapter 5 will discuss the conclusion and results for this study, and elaborate on future work.

2

Airway Tract Models and Inverse Filtering

In this chapter we will introduce acoustic properties of tubes and how they are related to our modelling for snoring sound production. First, wave equation and its relation to acoustic theory will be demonstrated in section 2.1. Later, in section 2.1.6 and 2.1.7 we will investigate the models for tube endings and demonstrate how they effect the linear prediction method. Section 2.2 will introduce physical phenomena that is discarded in simple tube modelling. Later, we will introduce two models for modelling the snore sound production 2.3. In section 2.4 we will link our model to linear prediction and lattice filters. The last section 2.5 will be devoted to inverse filtering methods previously used to model glottal flow in speech production.

2.1. Acoustic Transmission Line Model

In order to explain the theory behind the acoustic transmission line model, we have to start with the traveling wave equation. Wave equations are represented by partial differential equations that represent processes varying with time and space. Wave Equation is a simplification of Navier-Stokes equation for fluids. The simplification assumes that the medium is linear and retains only first order terms linear in pressure. Linear medium guarantees that waves of different frequencies do not interact with each other, which holds for pressure levels below 140 dB[40]. This property is called *superposition* and will be used as n presumption for the following chapters as natural sounds, produced by human tissues do not exceed 100 dB.

Wave equations for liquid flow are shown in Equation 2.4 on the following page. This equation can be obtained from three properties: State Equation 2.1, which is the application of Hooke's law to fluids; Newton's mass law 2.2, which takes the pressure difference as the force that accelerates the medium inside the volume, and mass preservation law 2.3 which states that what flows in and does not come out again will remain inside the volume since mass does not vanish [15].

$$\partial\rho = \kappa\rho\partial p \tag{2.1}$$

$$\rho\frac{\partial u(x,y,z,t)}{\partial t} = \nabla p(x,t) \tag{2.2}$$

$$\frac{1}{\rho c^2}\frac{\partial p(x,t)}{\partial t} = \nabla \bullet u(x,y,z,t) \tag{2.3}$$

where ∇ indicates the gradient and $\nabla\bullet$ the divergence and

$p(x,t)$ = pressure at location x, time t

$u(x,y,z,t)$ = vector velocity of an air particle at location(x,y,z)

c = speed of sound in air

ρ = density of air inside the vocal tract

κ = volume compressibility

By combining two equations in 2.3 on the preceding page and 2.2 on the previous page, we can obtain the general wave equation in 2.4

$$\Delta^2\Psi(t) - \frac{1}{c^2}\frac{\partial^2\Psi(t)}{\partial t^2} = 0 \tag{2.4}$$

Ψ is a scalar quantity called *velocity potential*. Despite it has no physical significance, it's negative gradient is the *field velocity*. Hence, this representation is used to derive pressure and volume velocity equations. If we use complex notation 2.5 and assume complex exponential for time dependence and rewrite Equation 2.4, we obtain a simpler wave equation 2.6 which is a second order partial differential equation in a simpler form of Helmholtz Equation.

$$\Psi(t) = \Psi e^{j\omega t} \tag{2.5}$$

$$\Delta^2\Psi(\omega) - k^2\Psi(\omega) = 0 \tag{2.6}$$

where k is ω/c *wave number* or *phase coefficient* in m^{-1}. Δ^2 is the Laplacian operator that can be represented as Equation 2.7 in three dimensions for Cartesian coordinate system. Solving the wave equation 2.4 with D'Alambert's method we obtain the harmonic relation equation and obtain $k = \frac{\omega}{c} = \omega\sqrt{\rho}\kappa$.

$$\Delta^2\Psi = \frac{\partial^2\Psi}{x^2} + \frac{\partial^2\Psi}{y^2} + \frac{\partial^2\Psi}{z^2} \tag{2.7}$$

Due to superposition property, ω will be discarded in the forthcoming notations. Solution of the wave equation can be evaluated in different coordinate systems, which are Cartesian, Cylindrical and Spherical. Spherical solutions constitute of Bessel and Henkel functions. However, we will only be interested in Cartesian Coordinate system, as our interest will be on one direction and the general solution for three dimensional systems can be obtained by superimposing the one dimensional solution. As we restrict ourselves to the one dimensional domain, we assume that the diameter of the tubes that we are interested in are much smaller compared to the wavelength ($d \ll \lambda$) which guarantees axial propagation. By investigating the dimensions, we find:

$$\lambda_{4kHz} = \frac{c}{F} = \frac{340\mathrm{m\,s^{-1}}}{4000\mathrm{Hz}} = 8.5\mathrm{cm} \tag{2.8}$$

while diameter of the vocal tract is around 2-4 centimeters [60] [28]. This satisfies the property that vocal tract tubes are much smaller than the compared wavelength.

One-dimensional propagation also means that all wave planes perpendicular to the axial direction are planes of constant phase. Additionally, velocity is independent of location (y,z). Here, we also define the *volume velocity* as:

$$v(x,t) = S(x,t)\,u(x,t)$$

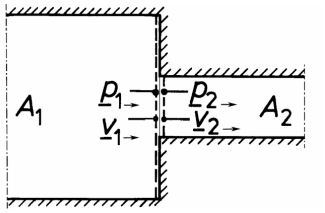


Figure 2.1: Tube with steps in the area function. Taken from [15]

where $S(x,t)$ is the cross sectional area of the vocal tract. For airway geometry, the area function is a non-constant function which changes with x, depth. The change between each cross section will be assumed as a step difference, which can be seen in figure 2.1. If we substitute volume velocity in equations 2.3 on page 7 and 2.2 on page 7 we get the following results

$$-\frac{u(x,t)}{\partial x} = \frac{1}{\rho c^2}\frac{\partial[p(x,t)S(x,t)]}{\partial t} + \frac{\partial S(x,t)}{\partial t} \tag{2.9}$$

$$-\frac{\partial p(x,t)}{\partial x} = \rho\frac{\partial[u(x,t)/S(x,t)]}{\partial t} \tag{2.10}$$

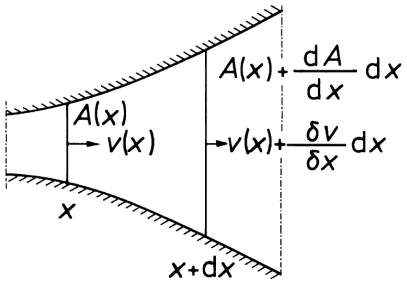


Figure 2.2: Cross-section of a horn, i.e. duct with gradually increasing cross area. Taken from [15]

To evaluate the wave equation, we first define the boundary conditions. Considering the elementary volume between the cross areas at x and $x+\mathrm{d}x$ in Figure 2.2 we return back to Euler's continuity equation 2.9 which states the inflowing and outflowing mass should be the same within a lossless tube.

$S(x)+\frac{\mathrm{d}S(x)}{2}$ is the average area of the horn. If we neglect the second-order differentials, we obtain 2.11,

$$-\rho\left(\frac{\partial u}{\partial x} + \frac{1}{S(x)}\frac{\mathrm{d}S}{\mathrm{d}x}u\right) = \frac{\partial \rho}{\partial t} \tag{2.11}$$

which is modified continuity equation. Using Equation 2.1 on page 7 we rewrite to get field quantity instead of ρ, and later combining with state equation we obtain Webster's Equation 2.13 that is used for conical and exponential horns.

$$-\left(\frac{\partial u}{\partial x} + \frac{1}{S(x)}\frac{\mathrm{d}S}{\mathrm{d}x}u\right) = \kappa\frac{\partial p}{\partial t} \tag{2.12}$$

It is complementary to remind here that term $\frac{1}{S(x)}\frac{\mathrm{d}S}{\mathrm{d}x}$ is identical to $\mathrm{d}[\ln S(x)/\mathrm{d}x]$. As it can be seen for constant area function, the equation reduces to one-dimensional wave equation 2.4 on the facing page

$$\frac{\partial^2 p}{\partial x^2} + \frac{1}{S(x)}\frac{\mathrm{d}S}{\mathrm{d}x}\frac{\partial p}{\partial x} = \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} \tag{2.13}$$

2.1.1. Characteristic Impedance

Here, we introduce the quantity characteristic impedance, which is sound pressure divided by particle velocity. Using Euler's equation 2.9 on the previous page and considering only forward progressing wave ($p(x) = \overrightarrow{p}\, e^{-jkx}$ and $u(x) = \overrightarrow{u}\, e^{-jkx}$) we simplify the division as follows.

$$Z_w = \frac{\overrightarrow{p}\, e^{-jkx}}{\overrightarrow{u}\, e^{-jkx}} = -\frac{\overleftarrow{p}\, e^{-jkx}}{\overleftarrow{u}\, e^{-jkx}} = \rho c \tag{2.14}$$

Z_w is real in no damping case, since sound pressure and particle velocity are in phase, which generally holds for plane waves in lossless fluids. If damping would be included in the equation k the wave number can be used as a complex number[40].

$$k = \frac{\omega}{c}(1 + j\alpha) \tag{2.15}$$

where α being the *damping coefficient*. However, for lossless tube model $\alpha = 0$. Damping effect can also be represented by resistance elements in the transmission line model. This will be discussed in section 2.2.1 on page 19.

2.1.2. Transmission-Line Equations and Reflectance

We aim to find the transmission line equations for a single acoustic duct with non-varying area in this section portrayed in Figure 2.3 on the next page. Hence, we take $\partial S(x,t)/\partial t = 0$ and $S(x,t) = S$ and replace in Webster's Equation 2.13 on the preceding page. Equations 2.9 on the previous page and 2.10 on the preceding page yields

$$-\frac{v(x,t)}{\partial x} = \frac{S}{\rho c^2}\frac{\partial[p(x,t)]}{\partial t}$$
$$-\frac{\partial p(x,t)}{\partial x} = \frac{\rho}{S}\frac{\partial[v(x,t)]}{\partial t}$$

As these equations are linear constant coefficient differential equations, the steady-state solution will have the form in Equation 2.16 which represents waves moving in both the positive and negative x directions

$$v_n(x_n,t) = \overrightarrow{v}_{(}x_n,t) + \overleftarrow{v}_{(}x_n,t) \tag{2.16}$$
$$= \frac{S_n}{\rho c}[\Psi_n^+(\omega)e^{j\omega t}e^{-j\omega(x_n/c)} - \Psi_n^-(\omega)e^{j\omega t}e^{+j\omega(x_n/c)}] \tag{2.17}$$
$$p_n(x_n,t) = p_0 = \overrightarrow{p}\,(x_n,t) + \overleftarrow{p}_{(}x_n,t) \tag{2.18}$$
$$= \Psi_n^+(\omega)e^{j\omega t}e^{-j\omega(x_n/c)} + \Psi_n^-(\omega)e^{j\omega t}e^{+j\omega(x_n/c)} \tag{2.19}$$

x_n representing displacement along the tube n (i.e. $0 \le x_n \le l_n$).

We also define *delay* as the time difference between the beginning of a tube and the end of a tube.

$$\tau_n = \frac{l}{c} \tag{2.20}$$

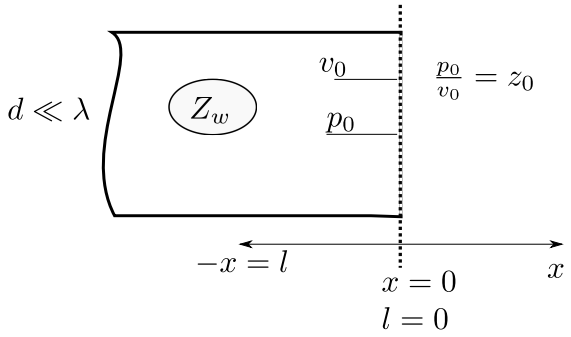


Figure 2.3: Simple Tube Model at the boundary

2.1.3. Termination of Tubes

At this point, it is important to inspect the boundary conditions for an acoustic tube. For this section we will assume the tube is excited by an exponential volume velocity with

$$v(0,t) = \mathrm{v}_{\mathrm{glottis}}(t) = \mathrm{v}_{\mathrm{glottis}}(\omega)e^{j\omega t} \tag{2.21}$$

Additionally, for this section we will consider $n = 1$ as we will inspect wave equations in a single tube.

Open Termination

For this case, the end of the tube is open, resulting in soft termination. We neglect the radiation effect at the end of the tube and take the boundary conditions as

$$p(l,t) = 0 \tag{2.22}$$
$$Z_l = 0 \tag{2.23}$$

Using Equation 2.21 and Equation 2.22 we obtain

$$\frac{S}{\rho c}\left[\Psi^+(\omega) - \Psi^-(\omega)\right] = \mathrm{v}_{\mathrm{glottis}} \tag{2.24}$$
$$\Psi_n^+(\omega)e^{-j\omega\tau} + \Psi_n^-(\omega)e^{+j\omega\tau} = 0 \tag{2.25}$$

Solving the equations for Ψ^+ and Ψ^-

$$\Psi_n^+(\omega) = \mathrm{v}_{\mathrm{glottis}}\frac{\rho c}{S}\left[\frac{e^{-j\omega\tau}}{2\cos(\omega\tau)}\right] \tag{2.26}$$
$$\Psi_n^-(\omega) = -\mathrm{v}_{\mathrm{glottis}}\frac{\rho c}{S}\left[\frac{e^{-j\omega\tau}}{2\cos(\omega\tau)}\right] \tag{2.27}$$

and from Equation 2.16 on the facing page and (2.18) steady-state solutions for volume velocity and pressure become

$$v(x,t) = \mathrm{v}_{\mathrm{glottis}}\left[\frac{\cos(\omega(\tau - \frac{x}{c}))}{\cos(\omega\tau)}\right]e^{-j\omega\tau} \tag{2.28}$$
$$p(x,t) = j\frac{\rho c}{S}\mathrm{v}_{\mathrm{glottis}}\left[\frac{\sin(\omega(\tau - \frac{x}{c}))}{\cos(\omega\tau)}\right]e^{-j\omega\tau} \tag{2.29}$$

As an example for position at $x = l$, at the boundary, the solution becomes

$$v(l,t) = \frac{\mathrm{V}_{\mathrm{glottis}}}{\cos(\omega l/c)}e^{j\omega t} = \mathrm{V}_{\mathrm{lips}}(\omega)e^{j\omega t} \tag{2.30}$$

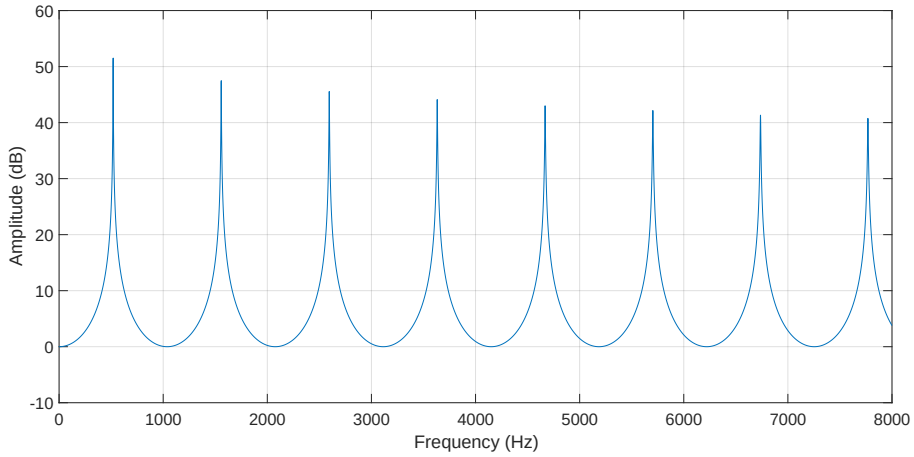


Figure 2.4: Tube Transfer Function for an open ended tube with 17.5 cm length

In turn the transfer function of vocal tract becomes

$$H(\omega) = \frac{1}{\cos(\omega l/c)}$$

The resonant frequencies can be found by setting the denominator equal to zero. Since $\omega_i = 2\pi F_i$ resonant frequencies occur at

$$F_i = \frac{c}{4l}[2i-1] \qquad for \quad i = 1,2,3,4... \tag{2.31}$$

As a result, transfer function for a tube of 17.5 cm and a speed of sound 343 m/sec resonant frequencies occur around 500,1500,2500 Hz as seen in Figure 2.4

Closed Termination

We derive the solutions for a closed termination which is at distance x in a lossless tube with length l. This representation has also been called as hard termination. It is achieved by closing the surface with rigid materials and the waves are fully reflected.

$$v(l,t) = 0$$
$$Z_l = \infty$$

Similarly to section 2.1.3.1 on the preceding page, open termination, we can compute pressure and volume velocity.

$$v(x,t) = \mathrm{v}_{\mathrm{glottis}}\left[\frac{\sin(\omega(\tau - \frac{x}{c}))}{\sin(\omega\tau)}\right]e^{-j\omega\tau}$$
$$p(x,t) = j\frac{\rho c}{S}\mathrm{v}_{\mathrm{glottis}}\left[\frac{\cos(\omega(\tau - \frac{x}{c}))}{\sin(\omega\tau)}\right]e^{-j\omega\tau}$$

Since the ratio of pressure and and volume velocity inside the tube is only dependent on the terminating impedance of the tube, $Z_0 = p_0/v_0$, it is convenient to formulate the solution of the wave equation in terms of the distance to the position of Z_0. To this end, we will substitute the coordinate x with $-l$ shown in Figure 2.3 on the previous page.

Now, because of $e^{-\frac{j\omega x}{c}} = 1$ at location $x = 0$, we have :

$$p_0 = \overrightarrow{p} + \overleftarrow{p} \tag{2.32}$$
$$u_0 = \overrightarrow{u} + \overleftarrow{u} \tag{2.33}$$

Keeping in mind Equation 2.14 on page 10 we can obtain:

$$Z_w v_0 = \overrightarrow{p} - \overleftarrow{p} \tag{2.34}$$

$$\frac{p_0}{Z_w} = \overrightarrow{u} - \overleftarrow{u} \tag{2.35}$$

Hence, we can follow with

$$\overrightarrow{p} = \frac{1}{2}\Big(p_0 + Z_w u_0\Big) \qquad \text{and} \qquad \leftarrow p = \frac{1}{2}\Big(p_0 - Z_w u_0\Big) \tag{2.36}$$

$$\rightarrow u = \frac{1}{2}\Big(\frac{p_0}{Z_w} + u_0\Big) \qquad \text{and} \qquad \leftarrow u = -\frac{1}{2}\Big(\frac{p_0}{Z_w} - u_0\Big) \tag{2.37}$$

For an arbitrary position $l = -x$ we get the following equations

$$p(l) = \frac{1}{2}\Big(p_0 + Z_w u_0\Big)e^{jkl} + \frac{1}{2}\Big(p_0 - Z_w u_0\Big)e^{-jkl} \tag{2.38}$$

$$v(l) = \frac{1}{2}\Big(\frac{p_0}{Z_w} + u_0\Big)e^{jkl} - \frac{1}{2}\Big(\frac{p_0}{Z_w} - u_0\Big)e^{-jkl} \tag{2.39}$$

Using Euler's identity $e^{j\alpha} = \cos(\alpha) - j\sin(\alpha)$ we obtain the transmission-line equations.

$$p(l) = p_0 \cos(kl) + jZ_w u_0 \sin(kl) \tag{2.40}$$

$$u(l) = j\frac{p_0}{Z_w}\sin(kl) + u_0 \cos(kl) \tag{2.41}$$

and T matrix for a specific duct becomes

$$\begin{bmatrix} p_1 \\ u_1 \end{bmatrix} = \begin{bmatrix} \cos(kl) & jZ_W \sin(kl) \\ j\frac{1}{Z_W}\sin(kl) & \cos(kl) \end{bmatrix} \begin{bmatrix} p_2 \\ u_2 \end{bmatrix} \tag{2.42}$$

2.1.4. Electrical Analogy

This T-matrix multiplication in Equation 2.42 has been used in our model with electrical circuit analogy, which can be represented by an L-C circuit. Here, it is essential to describe electro-acoustic analogies that are commonly used in representation of acoustic phenomena [17]. This analogy can be summarized in table 2.1 and table 2.2.

Table 2.1: Source Analogy for Acoustic - Electrical Circuits

Type		Mechanical Variable	Configuration
Power Conjugate Pair	Effort Variable	Pressure	Voltage
	Flow Variable	Volume Flow Rate	Current
Hamiltonian Variables	Effort Hamiltonian	Pressure-Momentum	Flux Linkage
	Flow Hamiltonian	Volume	Charge

Table 2.2: Impedance Analogy for Acoustic - Electrical Circuits

Acoustic Property	Analogous Electrical Property
Acoustic Resistance	Resistance
Acoustic Mass	Inductance
Acoustic Compliance	Capacitance
Acoustic Impedance	Electrical Impedance

These idealized equations mentioned in this section are different than a real physical medium. For these idealized equations to be used as they are, numerous conditions has to be satisfied:

1. The medium is homogeneous and does not crack, meaning that there are no inclusions of vacuum as might be caused by cavitation
2. Thermal conductivity is zero, which assumes adiabatic compression
3. Inner friction is zero meaning that there are no energy losses and no viscosity
4. The medium has defined mass and elasticity

Given that these conditions are satisfied, we can represent our tube model as a simple LC circuit. Two port theory says that [16]

$$Z_1 = j\frac{Z_w}{S}\tan\frac{kl}{2} \tag{2.43}$$

$$Z_2 = -j\frac{Z_w}{S}\frac{1}{\sin kl} \tag{2.44}$$

For sections of small lengths, that is $d \ll \lambda$, the following approximations apply

$$\tan\frac{kl}{2} = \frac{kl}{2} \quad and \quad \frac{1}{\sin kl} = \frac{1}{kl} \tag{2.45}$$

This can be seen in Figure 2.5. We concatenate multiple LC circuits to represent the vocal tract. This model corresponds to Kelly Lochbaum Lattice Filter, which will be mentioned in section 2.1.5.

$$L = \frac{klZ_w}{S} \tag{2.46}$$

$$C = \frac{Z_w}{klS} \tag{2.47}$$

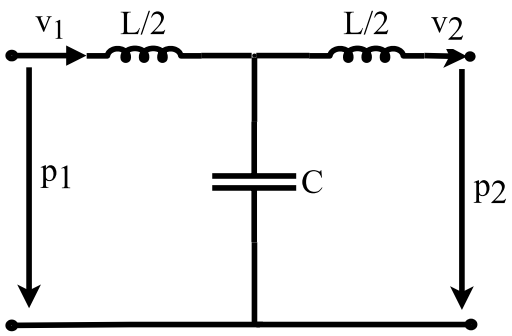


Figure 2.5: Simple Electrical equivalent T-Network circuits of vocal-tract section

2.1.5. Stepped Ducts and Lattice Filter

In order to determine the interaction of traveling waves between tubes, consider what happens at the juncture between the *k*th and *(k+1)*st tubes. Since sound wave propagation in a tube obeys the law of continuity, sound pressure and volume velocity must be continuous in both time and space everywhere along the tube model. Now, if we take boundary conditions as taking total volume velocity Sv, and pressure p as continuous at the the tube junction, i.e.

$$p_1 = p_2 \tag{2.48}$$

$$v_1 = S_1 u_1 = S_2 u_2 = v_2 \tag{2.49}$$

we get a reflected wave. As in characteristic impedance equation 2.14 on page 10 we obtain equation below from Equation 2.48

$$p(x,t) = \overrightarrow{p_1} + \overleftarrow{p_1} = \overrightarrow{p_2} \tag{2.50}$$

$$S_1\left(\frac{\overrightarrow{p_1}}{Z_w} - \frac{\overleftarrow{p_1}}{Z_w}\right) = S_2\frac{\overrightarrow{p_2}}{Z_w} \tag{2.51}$$

we define reflectance as proportion of pressure of backward propagating wave to pressure of forward propagating wave. As a result, we get Equation 2.52

$$r = \frac{\overleftarrow{p_1}}{\overrightarrow{p_2}} = \frac{S_1 - S_2}{S_1 + S_2} \tag{2.52}$$

As our characteristic impedance is defined as $Z_w = \overrightarrow{p} / \overleftarrow{p}$ we have to find the new tube equations. Now keeping in mind Equation 2.36 on page 13 we can find Z_L as

$$Z_L = \frac{Z_w}{S} = \frac{\rho c}{S} \tag{2.53}$$

As a result, T-matrix from equation 2.42 on page 13 becomes

$$\begin{bmatrix} p_1 \\ v_1 \end{bmatrix} = \begin{bmatrix} \cos(kl) & jZ_L\sin(kl) \\ j\frac{1}{Z_L}\sin(kl) & \cos(kl) \end{bmatrix} \begin{bmatrix} p_2 \\ v_2 \end{bmatrix} \tag{2.54}$$

Now keeping in mind Equation 2.48 on the preceding page we can incorporate *reflection coefficients* into our equation. It is easier and useful to model the tube with reflection coefficients. However, one has to keep in mind that reflection coefficients can model two tubes with a focus on the area change between two tubes while acoustic tube T-matrix in equation 2.42 on page 13 models a single tube. Hence, concatenating T-matrices gives the same result as the reflection coefficient model if the ideal conditions in section 2.1.4 on page 13 are satisfied. This hypotheses is tested in section 3.1 on page 27.

Thus from Equations 2.48 on the preceding page, 2.32 on page 12, 2.14 on page 10 we can rewrite the equations as transfer function between the volume velocities

$$\overrightarrow{v_1} + \overleftarrow{v_1} = \overrightarrow{v_2} + \overleftarrow{v_2} \tag{2.55}$$

$$Z_L(\overrightarrow{v_1} - \overleftarrow{v_1}) = Z_L(\overrightarrow{v_2} - \overleftarrow{v_2}) \tag{2.56}$$

From here, with a algebraic manipulation and using characteristic impedance equation 2.14 on page 10 we can get

$$\overrightarrow{v_2} = \frac{2S_2}{S_1 + S_2}\overrightarrow{v_1} + \frac{S_2 - S_1}{S_1 + S_2}\overleftarrow{v_2} \tag{2.57}$$

$$\overleftarrow{v_1} = \frac{S_2 - S_1}{S_1 + S_2}\overrightarrow{v_1} + \frac{2S_2}{S_1 + S_2}\overleftarrow{v_2} \tag{2.58}$$

From this equation we make an important physical assumption by equating the negative travelling wave in the second tube to zero. We obtain the following relation.

$$\overrightarrow{v_2} = \frac{2S_2}{S_1 + S_2}\overrightarrow{v_1} \tag{2.59}$$

$$\overleftarrow{v_1} = \frac{S_2 - S_1}{S_1 + S_2}\overrightarrow{v_1} \tag{2.60}$$

Here we can define *transmission coefficients* and *reflection coefficients* , respectively, as

$$\Gamma_n^+ = \frac{\overrightarrow{v_2}}{\overrightarrow{v_1}} = \frac{2S_2}{S_1 + S_2} \tag{2.61}$$

$$\Gamma_n^- = \frac{\overleftarrow{v_1}}{\overrightarrow{v_1}} = \frac{S_1 - S_2}{S_1 + S_2} = \Gamma_n^+ - 1 \tag{2.62}$$

we take $\Gamma = \Gamma_n^-$ as the reflection coefficients. One has to keep in mind that area functions are always positive, reflection coefficients are bounded in magnitude by one.

$$-1 \leq \Gamma \leq 1$$

We now rewrite the Equations 2.57 on the previous page and 2.58 on the preceding page with reflection coefficient formulation in 2.62. Here, we also incorporate delay, τ in our formulation.

$$\overrightarrow{v_2}(t) = (1+\Gamma)\overrightarrow{v_1}(t-\tau) + (\Gamma)\overleftarrow{v_2}(t) \tag{2.63}$$

$$\overleftarrow{v_1}(t+\tau) = (-\Gamma)\overrightarrow{v_1}(t-\tau) + (1-\Gamma)\overleftarrow{v_2}(t) \tag{2.64}$$

These equations are called *Kelly-Lochbaum equations*. We can now represent our model with a signal flow graph for a single tube junction.

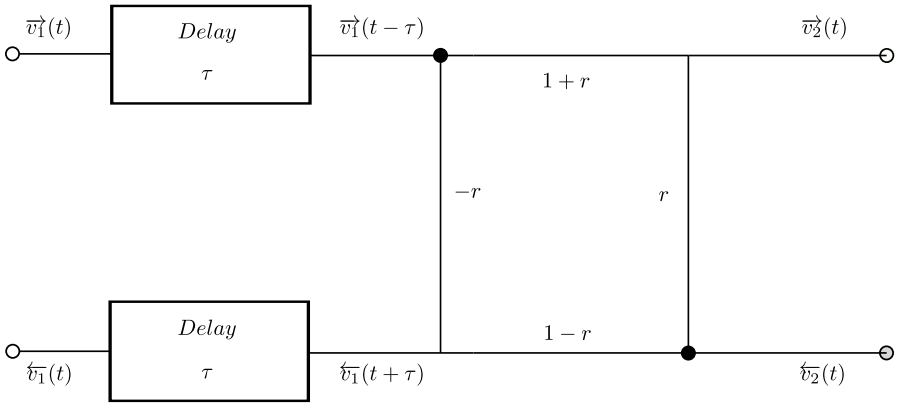


Figure 2.6: Kelly-Lochbaum Lattice Filter

and equivalently, taking z-transform for $z = e^{j\omega 2\tau}$

$$\overrightarrow{v_2}(z) = z^{-1}(1+\Gamma)\overrightarrow{v_1}(z) \quad + \quad \Gamma\overleftarrow{v_2}(z) \tag{2.65}$$

$$\overleftarrow{v_2}(z) = \frac{1}{1-\Gamma}\overrightarrow{v_1}(z) \quad + \quad \frac{\Gamma z^{-1}}{1-r}\overleftarrow{v_1}(z) \tag{2.66}$$

These equations are useful as we discover in section 2.4 on page 21 there are methods in signal processing to solve these equations with $O(n^2)$ complexity compared to $O(n^3)$ in Gaussian Elimination [46].

2.1.6. Lips

If we consider the junction between the last vocal tract tube and the area beyond the lips, we can say that the junction can be modeled by an infinite baffle. In ideal situations the sound wave radiates outward, which signifies that backward travelling sound wave, $v_{(N+1)\leftarrow}$ should be zero. Hence impedance for a half infinite tube with cross-sectional area S_{N+1} becomes

$$Z_{Lips} = \frac{\rho c}{S_{N+1}}$$

Accordingly, from backward travelling sound wave and Equation 2.65 on the preceding page we can find

$$\overleftarrow{v_N}(t+\tau) = -\Gamma_{lips}\overrightarrow{v_N}(t-\tau) \tag{2.67}$$

As an alternative approach, in many of the electrical models of the vocal tract, the radiation impedance of the mouth is represented by a serial connection of acoustic resistance and reactance. We need to approximate the radiation impedance to be relatively simple for mathematical calculation. Radiation resistance represents energy radiated away from the mouth. The reactance represents effective mass of the vibrating air between the lips.

However, we can create a better model for lip radiation. As air between the lips oscillates in the same way a piston does; the entire surface oscillates in phase. This is why radiation from a rigid circular piston can be taken as a basic model of mouth aperture's radiation. The ideal way to simulate a speaker's mouth is to use a circular piston set in a spherical baffle. However, as this solution is computationally expensive, it is beneficial to use a piston set in an infinite plane baffle. Equation for a piston set in an infinite baffle is

$$Z_p = \frac{\rho c}{S}\left[(1-\frac{J_1(2kr)}{kr}) + j\frac{H_1(2kr)}{kr}\right] \tag{2.68}$$

with wavelength k, radius r. $J_1(x)$ is the Bessel Function of the first kind and $H_1(x)$ is the Struve function. Here we use five different approximations. First one being, Fant's estimation with infinite baffle [66]

$$Z = \frac{\rho c}{S}\left[\frac{(kr)^2}{2} + j\frac{8kr}{3\pi}\right] \tag{2.69}$$

and equates to

$$R = 128\frac{\rho c}{9\pi^2 S} \tag{2.70}$$

$$L = 8\frac{\rho}{3\pi\sqrt{\pi S}} \tag{2.71}$$

The Fant model in Equation 2.69 can be improved by taking one more term of the infinite series, thus increasing the accuracy of the correction functions in Chalker et al.[24]. In that case, the radiation impedance of the mouth aperture is given by the expression

$$Z = \frac{\rho c}{S}\left\{\frac{(kr)^2}{2}\left[1-\frac{(kr)^2}{6}\right] + j\frac{8kr}{3\pi}\left[1-\frac{2(kr)^2}{15}\right]\right\} \tag{2.72}$$

and lastly we used Wakita and Fant's proposed expression for radiation impedance

$$K(\omega) = \begin{cases} 1.6; & (w \geq 2\pi 1600) \\ \frac{0.6\omega}{(2\pi 1600)} + 1; & (0 \leq \omega \leq 2\pi 1600) \end{cases} \tag{2.73}$$

$$Z = \frac{\rho c}{S}\left[\frac{(kr)^2}{4}K(\omega) + j0.8kr\right] \tag{2.74}$$

Lastly, we use an Approximation of the Struve function H_1 occurring in impedance calculations from Aarts et al.[1]

$$Z = \rho c\pi r^2\left[1-\frac{2J_1(2kr)}{2kr} + 2j\frac{H_1(2kr)}{2kr}\right]; \tag{2.75}$$

$$H_1(z) = \frac{2}{\pi} - J_0(z) + \left(\frac{16}{\pi}-5\right)\frac{\sin z}{z} + \left(12-\frac{36}{\pi}\right)\frac{1-\cos z}{z^2} \tag{2.76}$$

The results can be seen in Figure 2.7 on the next page

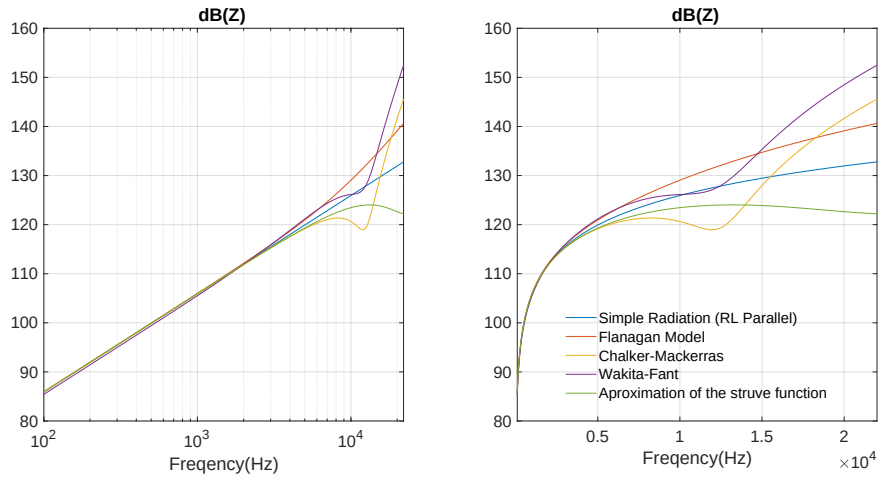


Figure 2.7: Lip Radiation in Linear and Logarithmic Scale

It can be seen that lip radiation models have a very small impact under frequencies below 10.000 Hertz. This means that the models that will be chosen will have a negligible impact for the lower frequency range, which we are interested in.

2.1.7. Glottis and Lower Respiratory Tract

Glottis can also be modeled in different ways. A simpler model, that is used in LP analysis is a resistance and impedance connected parallel to the current source, i.e. volume velocity. If we use a first order linear approximation for impedance we can use a time-invariant impedance as

$$Z_{glottis}(\omega) = R_{glottis} + j\omega L_{glottis}$$

This works as a current divider, and the forward traveling sound wave at the boundary $\overrightarrow{v_1}$ becomes

$$\overrightarrow{v_1} = \frac{1+\Gamma_{glottis}}{2} u_{\text{glottis}}(t) + \Gamma_{glottis}\overleftarrow{v_1} \tag{2.77}$$

where glottal reflection coefficient is

$$\Gamma_{glottis} = \frac{Z_{glottis} - \rho c/S}{Z_{glottis} + \rho c/S} \tag{2.78}$$

In z-transfer domain we find

$$V_{\text{glottis}}(z) = \begin{bmatrix} \frac{2}{1+r_{glottis}} & \frac{-2\Gamma_{glottis}}{1+\Gamma_{glottis}} \end{bmatrix} \begin{bmatrix} \overrightarrow{v_1} \\ \overleftarrow{v_1} \end{bmatrix} \tag{2.79}$$

This glottis model has been used in linear prediction model in Section 2.4 on page 21.

2.2. Additional Parameters

In section 2.1.4 on page 13 it has been mentioned that the lattice filter can only represent a T-network circuit with a capacitance and inductance which correspond to acoustic inertance and acoustic compliance. However, as mentioned by Flanagan [38] and Ishizaka [48], vocal tract can be better modeled by additional parameters. This section will be devoted to give a comprehensive understanding of how these parameters impact the inverse filter reconstruction with linear prediction coefficients. Later, the impact on transfer function of the vocal tract and cross-sectional area reconstruction with linear prediction will be discussed in

Chapter 3 on page 27. The additional parameters that will be investigated in this section can be seen in Figure 2.8

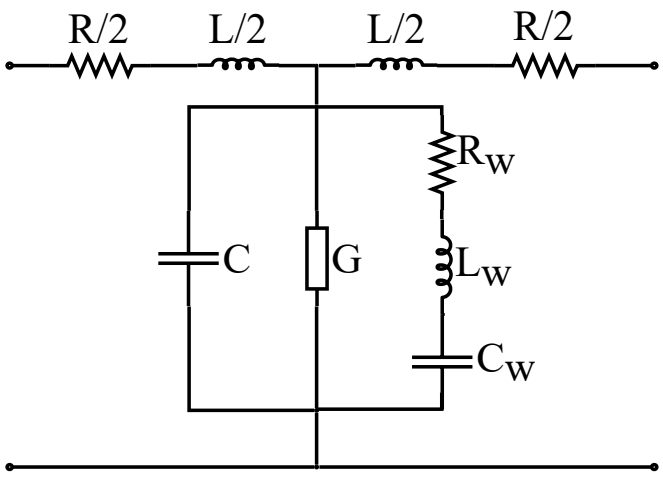


Figure 2.8: Electrical equivalent T-Network circuits of vocal-tract section

Parameter	SI unit	Value
air density ρ	$\mathrm{kg\,m^{-3}}$	1.14
dynamic viscosity μ	Pa s	1.85×10^{-5}
thermal conductivity λ	$\mathrm{W\,m^{-1}\,K^{-1}}$	0.0263
adiabatic compression / ratio specific heats γ	-	1.402
speed of sound c	$\mathrm{m\,s^{-1}}$	353

Table 2.3: Parameters for the physical properties of air in upper airway [79]

2.2.1. Resistance

The acoustic resistance R represents a power loss proportional to volume velocity and is the power dissipated in viscous friction at the tube wall. The easiest way to put in evidence this equivalent surface resistance is to consider the situation in which the tube wall is a plane surface, and moving sinusoidally in the x-direction with a velocity. The air particles that flow nearby the wall are under effect of a force caused by the shear viscosity, μ, of the medium. The power that was dissipated per unit length, l, can be calculated by determining the total power dragging the air with the plate. Detailed explanation of the Equation 2.80 can be found in Flanagan [38].

$$R = \frac{2}{\pi r^3}\sqrt{\frac{\rho\mu\omega}{2}} \tag{2.80}$$

where r is the radius of the tube, ω the angular frequency, ρ the density of air.

2.2.2. Wall Parameters

Up until now it has been assumed that the walls of the airway tract to be smooth and rigid. The dissipative elements of concern are then the radiation resistance, the glottal resistance, and the viscous and heat conduction losses at the cavity walls. In reality, vocal tract is not hard-walled, and it's surface impedance is finite. We aim to estimate this effect using previous research on wall parameters.

As our medium is not ideal we include wall parameters to our equation in parallel. These parameters have been used in [66] [82]. It is suggested that these wall parameters are connected as serially connected to each other and parallel to the lumped model as seen in Figure 2.8 on the previous page. There are no values that are agreed upon wall parameters. Ishizaka et al [48] measured the mechanical impedance of the cheek under relaxed and tensed conditions. Fant et al. [48] measured impedance using resonance characteristics of the vocal tract under the closed tract condition and proposed using lumped wall impedance, independent of vocal tract configurations. In section 3.2 on page 29 response for these additional parameters are investigated.

One has to keep in mind that in Ishizaka et al. [48] the calculation is done for per-unit wall areas. Hence, the impedance formula becomes

$$Z_w = \frac{B + j(\omega M - K/\omega)}{2\pi r l} \tag{2.81}$$

The other formulation for wall impedance is given in Harper et al.[44]. The measurements for the tissue have been investigated in Habib et al. [43], Capper et al. [22]. As a result impedance becomes

$$Z_w(\omega) = \frac{h}{2\pi r^3}\left[\nu(\omega) + j\omega\left[\rho r^2 - \frac{E(\omega)}{\omega^2}\right]\right] \tag{2.82}$$

Parameter	Value	Units
Resistance	$R_w(\omega) = \frac{h\nu(\omega)}{2\pi r^3 l}$	$\frac{kg}{m^4 s}$
Inertance	$L_w(\omega) = \frac{h\rho}{2\pi r l}$	$\frac{kg}{m^4}$
Compliance	$C_w(\omega) = \frac{2\pi r^3 l}{hE_w(\omega)}$	$\frac{m^4 s^2}{kg}$

Table 2.4: Nonrigid Wall , Lumped Parameters For A Segment Of Length l

2.2.3. Conductance

The analogous shunt conductance provides a power loss proportional to the square of the local sound pressure. Such a loss arises from heat conduction at the walls of the tube. The per-unit-length conductance can be deduced in a manner similar to that for the viscous loss. The derivation can be found in [38].

$$G = \frac{2\pi r l(\gamma - 1)}{\rho c^2}\sqrt{\frac{\lambda w}{2c_p \rho}} \tag{2.83}$$

2.3. Snoring Models

In this section, we will define the two models that we suggest to model the acoustic vocal tract, with an excitation located within the tube.

2.3.1. Model 1: Anechoic Termination at Excitation Location

This model suggests that at the location of the excitation signal, the tube is terminated with a characteristic impedance of $\rho c/S_E$, where E being the depth of the excitation location. As we assume that the tube ends at the excitation location, the linear prediction or Levinson-Durbin recursion finds that the rest of the tube has $\Gamma = 0$, i.e.

$$\Gamma_j = 0 \quad for \quad j = E+1, E+2..., N$$

This model is assumed to be the supporting model for inverse filtering methods in section 2.5 on page 24. It shall be noted that, many of the excitation locations are in the middle of the

airway tract for snoring sound production. For that reason, the order for linear prediction method in inverse filtering has to be halved.

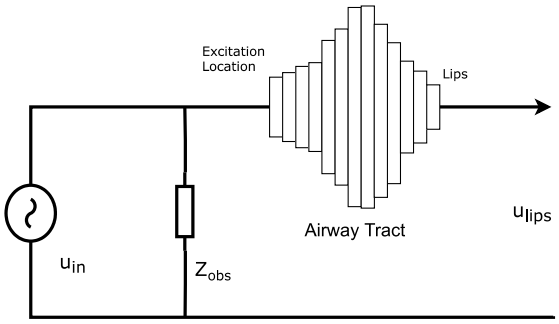


Figure 2.9: Circuit Diagram for Model 1

2.3.2. Model 2: Anechoic Termination At Glottis

In this model, we say that excitation location is independent of the tube ending. We move our current source within the tube, without changing the tube length. As a result, the schematic can be seen in figure 2.10. The lower part of the respiratory tract works as a current divider. We calculate the input impedance of the lower tract and add this impedance in parallel to upper part of the respiratory tract. Again, we have a current source, that is analogous to volume velocity source.

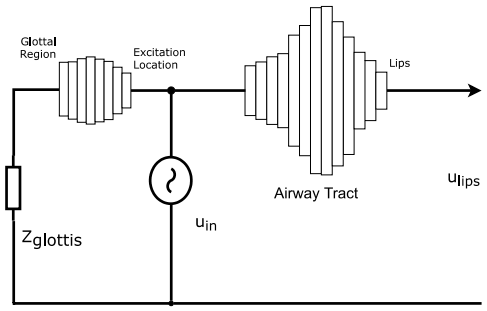


Figure 2.10: Circuit Diagram for Model 2

These models will be investigated in the Experiments and Results chapter.

2.4. Linear Prediction Analysis Method

The preceding lattice filter in section Equation 2.1.5 on page 14 has a close relation with linear prediction analysis. We can take Equation 2.65 on page 16 and concatenate N tubes. If we assume that last correlation coefficient, Γ_N, as the junction between the lips and the last tube we can find the lip radiation relation given in 2.67 on page 17 in z-domain as

$$V_{N+1}(z) = \begin{bmatrix} \overrightarrow{V_{lip}} \\ \overleftarrow{V_{lip}} \end{bmatrix} z^{-1/2} = \begin{bmatrix} \overrightarrow{V_{lip}} \\ 0 \end{bmatrix} \tag{2.84}$$

And Kelly-Lochbaum equations for N tubes become

$$\begin{bmatrix} \overrightarrow{V_k} \\ \overleftarrow{V_k} \end{bmatrix} = \frac{z}{1+\Gamma_k} \begin{bmatrix} 1 & -\Gamma_k \\ -\Gamma_k z^{-1} & z^{-1} \end{bmatrix} \begin{bmatrix} \overrightarrow{V_{k+1}} \\ \overleftarrow{V_{k+1}} \end{bmatrix} \tag{2.85}$$

Combining Equations 2.79 on page 18, 2.84 on the previous page, and 2.85, transfer function becomes

$$H(z) = \frac{\mathsf{V}_{\mathsf{lips}}(\mathsf{z})}{\mathsf{V}_{\mathsf{glottis}}(\mathsf{z})} = \frac{1+\Gamma_{glottis}}{2} \frac{z^{-N/2} \prod_{k=1}^{N} 1+\Gamma_k}{\begin{bmatrix} 1 & -\Gamma_{glottis} \end{bmatrix} \prod_{k=1}^{N} \begin{bmatrix} 1 & -\Gamma_k \\ -\Gamma_k z^{-1} & z^{-1} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix}} \tag{2.86}$$

As it can be seen, the transfer function is an all-pole model, that can be modeled by Kelly-Lochbaum lattice filter. Moreover, as a result, if we expand the denominator we find the Autoregressive (AR) filter as below

$$D(z) = 1 - \sum_{k=1}^{N} a_k z^{-1} \tag{2.87}$$

The Levinson-Durbin recursion is an algorithm for finding an all-pole IIR filter with a prescribed deterministic autocorrelation sequence. The filter that Levinson produces is minimum phase which means that all of the poles and zeroes are inside the unit circle. The filter, in turn, becomes a stable and causal filter. This property is also useful as the inverse, $1/H(z)$, is also stable and causal.

Levinson filter finds the coefficients of a $length(r)-1$ order AR linear process which has r as its a real or complex deterministic autocorrelation sequence. The denominator polynomial A(z); that is, $a = [1 \quad a(2) \quad \ldots \quad a(n+1)]$. The filter coefficients are ordered in descending powers of z^{-1}. We can also calculate reflection coefficients, Γ with a step-down recursion.

input : autocorrelation values r of length p
output: reflection coefficients Γ and coefficients of the all-pole model $a(k)$

1 $a_0(0) = 1$;
2 $\epsilon_0 = r_x(0)$;
3 **for** *j=0:p-1* **do**
4 $\quad \gamma_j = r_x(j+1) + \sum_{i=1}^{j} a_j(i) r_x(j-i+1)$;
5 $\quad \Gamma_{j+1} = \gamma_j / \epsilon_j$;
6 $\quad$ **for** *i=1:j* **do**
7 $\quad\quad a_{j+1}(i) = a_j(i) + \Gamma_{j+1} a_j^*(j-i+1)$;
8 $\quad$ **end**
9 $\quad a_{j+1}(j+1) = \Gamma_{j+1}$;
10 $\quad \epsilon_{j+1} = \epsilon_j [1 - |\gamma_{j+1}|^2]$;
11 **end**

Algorithm 1: Levinson-Durbin Recursion

As it can be seen in algorithm at line 7, relation between reflection coefficients and filter coefficients are coupled. If we flip the filter coefficients vector and take the reciprocal, we

find the following

$$\mathbf{a_j^R} = \begin{bmatrix} a_j^*(j) \\ a_j^*(j-1) \\ a_j^*(j-2) \\ \vdots \\ a_j^*(1) \\ 1 \end{bmatrix} \tag{2.88}$$

$$a_j^R(i) = a_j^*(j-i) \tag{2.89}$$

$$A_j^R(z) = z^{-j} A_j^*(1/z^*) \tag{2.90}$$

now, we can rewrite line 7 on the facing page in Z domain as

$$a_{j+1}(i) = a_j(i) + \Gamma_{j+1} a_j^R(i-1) \tag{2.91}$$

$$A_{j+1}(z) = A_j(z) + z^{-1}\Gamma_{j+1} A_j^R(z) \tag{2.92}$$

and conversely, if we take complex conjugate of both sides, we find

$$A_{j+1}^R(z) = z^{-1} A_j^R(z) + \Gamma_{j+1}^* A_j(z) \tag{2.93}$$

Combining Equation 2.92 and Equation 2.93 results in an almost identical relation between forward and backward going waves in Equation 2.85 on the preceding page with an exception for some scaling and a delay.

$$\begin{bmatrix} A_{j+1}(z) \\ A_{j+1}^R(z) \end{bmatrix} = \begin{bmatrix} 1 & z^{-1}\Gamma_{j+1} \\ \Gamma_{j+1}^* & z^{-1} \end{bmatrix} \begin{bmatrix} A_j(z) \\ A_j^R(z) \end{bmatrix} \tag{2.94}$$

As these relations are identical for the acoustic tube and lattice sections, given the following boundary conditions are satisfied, the reflection coefficients and cross sectional areas can be found using equation 2.62 on page 16. There can be two cases to introduce boundary conditions:

1. *Case 1:* Front-end of the tube is short circuited(open termination)($\Gamma_M = 1$) and back end of the tube is terminated with anechoic ending or wave match($1 < \Gamma_{glottis} < 0$). This means that impedance of the ending is the characteristic impedance of the last tube. The tube is excited by a volume velocity.

2. *Case 2:* Front end of the tube is terminated with anechoic ending($1 < \Gamma_{Glottis} < 0$) and back end is completely closed($\Gamma_{glottis} = 1$). Tube is excited by forward going volume velocity at the back end.

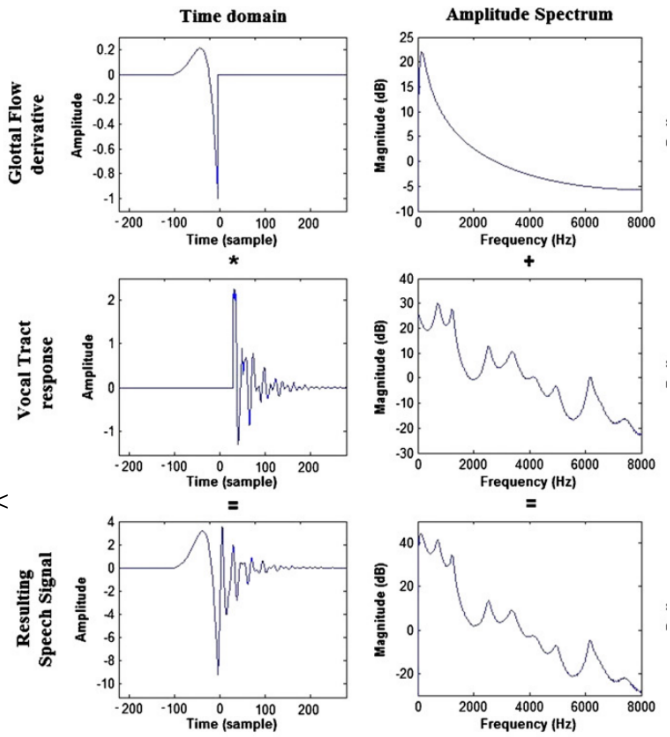


Figure 2.11: Illustration of source filter model. Rows respectively exhibit the following signals: the glottal flow derivative (top), the vocal tract response (middle), and the resulting speech signal (bottom). Columns represent time domain (first column), amplitude spectrum (second column). Image taken from [31]

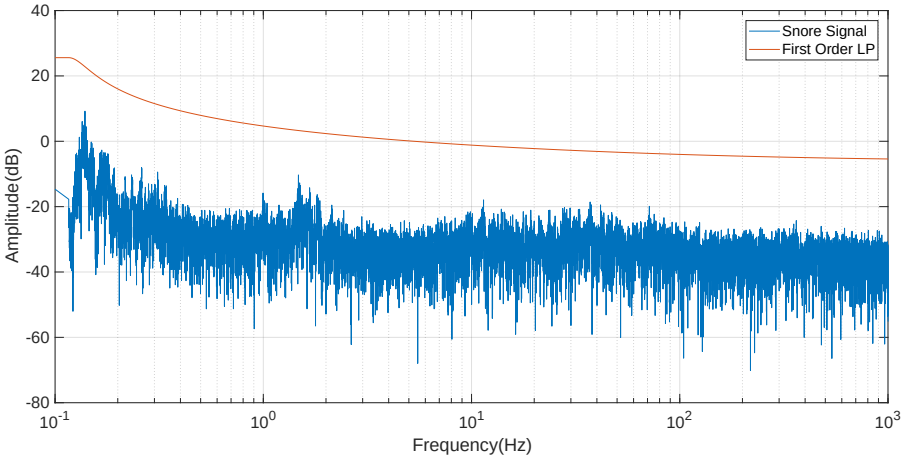


Figure 2.12: Comparison of Snore Excitation and Speech Glottal Flow Spectra

For Linear Prediction Analysis to estimate the cross-sectional areas, we have to implement the Case 1 [78].

2.5. Inverse Filtering

In our models, we assume that we have a biomechanical oscillator at the location of obstruction and acts as the excitation source. Similar to speech production theory, source flow will be called excitation flow in the following sections. In order to find the excitation flow, we will ground our model to source filter theory. Source filter theory assumes that source and the airway are linearly separable and excitation flow is filtered by airway transfer function and lip radiation transfer function. As mentioned in section 2.3.1 on page 20, theory of estimating cross sectional area of the airway using linear prediction assumes that ending of the tube is anechoic ending[78].

First, Direct Inverse Filtering (DIF) method will be mentioned. Then, Iterative Adaptive Inverse Filtering (IAIF) will be introduced. We will also discuss how an average waveform for flow models are found. Consequently, the results will be compared.

2.5.1. Direct Inverse Filtering

First, we filter the original signal to eliminate low-frequency fluctuations captured by the microphone. We will use a linear phase FIR filter with a cut-off frequency lower than the fundamental frequency of the snoring signal. For our model we used a model with passband frequency 60 Hz as we expect our fundamental frequency to be higher than 80 Hz for snoring signals.

Second, as lip radiation effect is assumed to be a derivative, we integrate the high-pass filtered signal to cancel this effect.

The glottal excitation, having a quasi-periodic structure, possesses a spectrum exhibiting a harmonic structure. The amplitudes of the harmonic components are traditionally considered to decrease monotonically at a rate of 12 dB/octave [35]. Here, we mention that the voice source for speech signals creates a spectrum slope of approximately -12dB/octave. It has been found that similar characteristics hold for snoring signals as seem in Figure 2.12. Later, we inverse filter this effect.

We now have a signal that represent the vocal tract filter excited with and impulse train. Here pth order model for the vocal tract is estimated. We then inverse filter this model and acquire a final estimate for glottal flow.

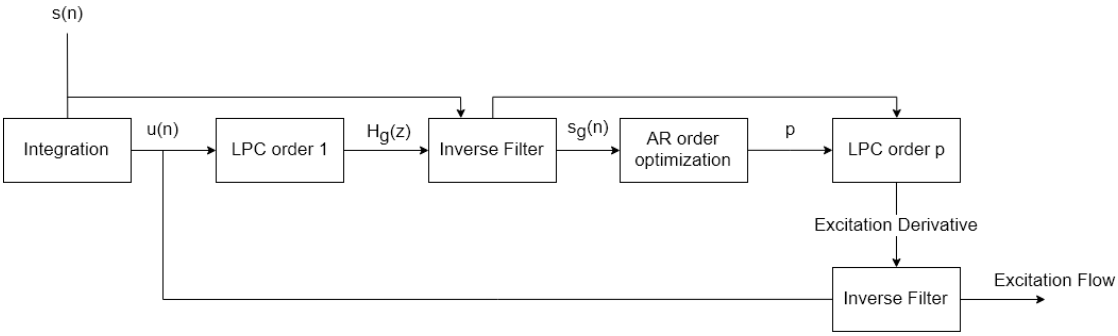


Figure 2.13: Direct Inverse Filter

2.5.2. Iterative Adaptive Inverse Filtering

Iterative Adaptive inverse filtering is an advanced method compared to direct inverse filtering[3]. Building blocks from 1 to 6 give the same result as direct inverse filtering. Hence blocks from 7 to 12 will be discussed in this section. At block seven, we estimate the contribution of excitation flow to snore spectrum, $H_{g2}(z)$ with a linear prediction order of 5 from first estimation of excitation flow. Then, we cancel this contribution from the input signal. This gives us a signal that constitutes Vocal tract and lip radiation. By integrating, lip radiation is removed in block 9. Finally, in block 10 a new vocal tract model, $H_{vt2}(z)$ is estimated with linear prediction.

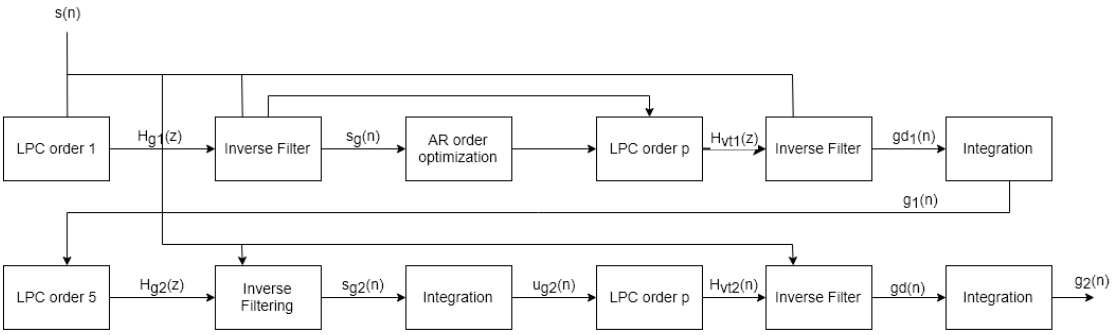


Figure 2.14: Iterative Adaptive Inverse Filter

3

Experiments and Results

This chapter will be devoted to explore the models suggested in Chapter 2 and analyze snoring sound signals from overnight recordings. First, in Section 3.2 impact of the additional parameters for airway tract model and reconstruction from the resulting transfer function with Linear Prediction analysis will be investigated. This is required to validate our model and investigate if the use of linear prediction is adequate for speech production and snoring sound production. In section 3.3, we will also investigate how changes in excitation location impacts the spectrum in both of the models mentioned in section 2.3 on page 20. Later, in section 3.4 we will validate our model by synthesizing vowels from inverse filtered real sound files, and convolving the resulting glottal flow with our spectrum from airway tract models. This section is essential to compare the transfer function from inverse filtering and the airway tract models. Last section 3.5 will be on analyzing and synthesizing real life snoring sounds. An estimation for excitation flow will be found using inverse filtering, and then will be used for synthesizing snoring sounds. We will also compare the two models introduced in section 2.3 on page 20 with spectrum of the transfer functions obtained from inverse filtering methods. Consequently, we will try to find a significant difference within the average excitation flows of different participants, which are impacted by different devices that alleviate the snoring problem.

For the default cross sectional area, we used the study of Mohsenin et al. [60] which is a study that measures differences of upper airway collapsibility based on gender. Subjects were encouraged to sleep on their backs (supine) as well as on their sides (non-supine) during the sleep study. The pharyngeal cross-sectional area was measured using an acoustic reflection technique. Previous study of Marshall et al. [57] shows that acoustically measured pharyngeal dimensions and those obtained through magnetic resonance imaging shows remarkable concordance. It shall be mentioned that the cross-sectional area is varying between patients, ages, BMI, gender, RDI and positions. We have taken the mean of the resting mandibular position among all patients. The results that we draw shall be taken cautiously as the cross sectional are changes based on the aforementioned factors.

Consequently, it is complementary to mention the findings of Dalmasso et al. [29] as the greater part of the energy content in snoring sounds is below 5,000 Hz and the main components lie in the low frequency range, at about 130 Hz and in the medium frequency range, at about 1000, 2200 and 3500 Hz. For our study we will also focus on lower frequencies.

3.1. Matching Models

In this section, our objective is to match reflection coefficient transfer matrix eq:TFreflection with concatenation of two ported T-network, namely tube model in Equation 2.42 on page 13 with an open end termination at the front end and terminated with a characteristic impedance of $\rho c/S_0$ at the back end. This is the case defined in Case 1 in section 2.4 on page 21. We will also demonstrate impact of $\Gamma_{glottis}$ for Case 1, as the value will be important for validating

the vocal tract transfer function in speech analysis in section 3.4.2 on page 38. We will later demonstrate the transfer response for second model mentioned in Section 2.3.2 on page 21.

3.1.1. Reflection coefficient of Glottis

Reflection coefficient of glottis, $\Gamma_{glottis}$, has a significant impact on the transfer function of the vocal tract in terms of amplitude of formant frequencies. However, as mentioned in Wakita et al. [78], reconstruction is not impacted by the changes in $\Gamma_{glottis}$ for

$$\{\Gamma \mid \Gamma_N \in \mathbb{Z}, 0 < \Gamma < 1\}$$

For the transfer function generation we used model 1, and did not include lip radiation or the additional parameters. Exclusion of lip radiation corresponds to open end termination, i.e. $\Gamma_M = 1$. We see that for varying reflection coefficient we have a very peaked transferred and for decreasing $\Gamma_{glottis}$, the peaks are rounded. Realistic values for the glottis parameters are assumed to be in the range of 0.6-0.8. We have taken 0.7 in the further analysis.

For the sampling frequency, we have taken 16000 Hz, same with our audio samples. If we define T as twice the time needed for a wave to propagate through a single section $T = \frac{2l}{c}$ and the relation between sampling frequency $f_s = 1/T$, number of sections and model order as P the relationship becomes.

$$P = \left\lceil \frac{2 f_s L}{c} \right\rceil \tag{3.1}$$

The order of the inverse filter has been determined by Equation 3.1, which corresponds to 15-16 orders for normal upper airway length (16cm-19cm). For reflection coefficients. reconstruction has been done assuming that area at the mouth is 1 and relationship between cross-sectional areas are found Using 2.62 on page 16. This can be observed in Figures 3.1 and 3.2 on the next page .

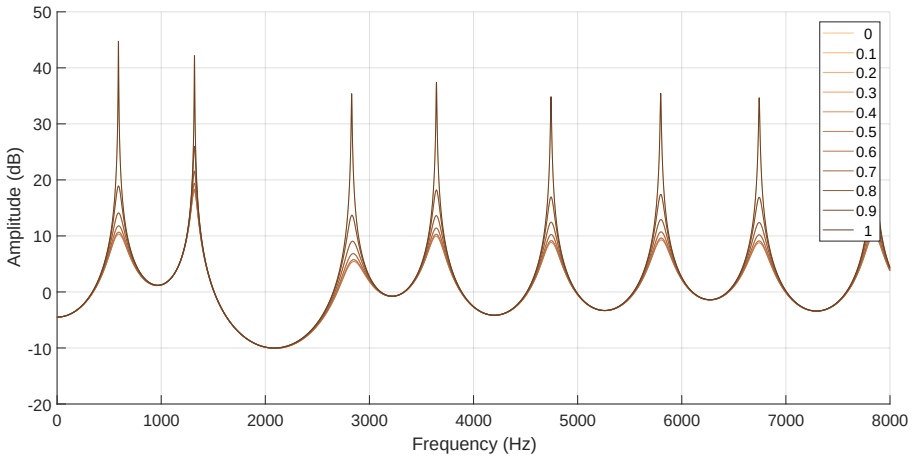


Figure 3.1: Transfer Function for changing $\Gamma_{glottis}$

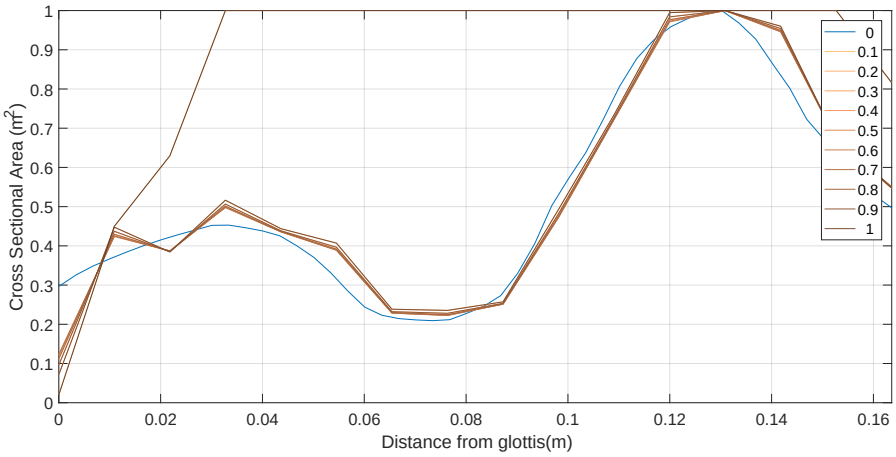


Figure 3.2: Effect of varying $\Gamma_{glottis}$ on reconstruction. Blue line is the cross-sectional area function, that is given as an input for tube model.

As it can be seen, reconstructed area of the airway tract closely follows the input area function with an exception of $\Gamma_{glottis} = 1 = \Gamma_{lips}$. This also shows that under ideal circumstances mentioned in 2.1.4 on page 13, linear prediction closely approximates the cross-sectional area.

3.2. Impact of Additional Parameters

In this section, we will investigate the additional parameters introduced in Section 2.2 on page 18. As mentioned in section 2.1.4 on page 13 these reflect physical properties of airway tract that can not be represented by a lattice filter. Our objective is to find how these additional parameters alter the reconstruction of cross-sectional area of the upper airway.

3.2.1. Resistance

As mentioned in 2.2.1 on page 19 resistance represents the power loss within the tube. We have used the values in Table 2.3 on page 19 to calculate the resistance. The transfer function and the reconstruction for given values can be found in 3.3 and 3.4 on the following page, respectively.

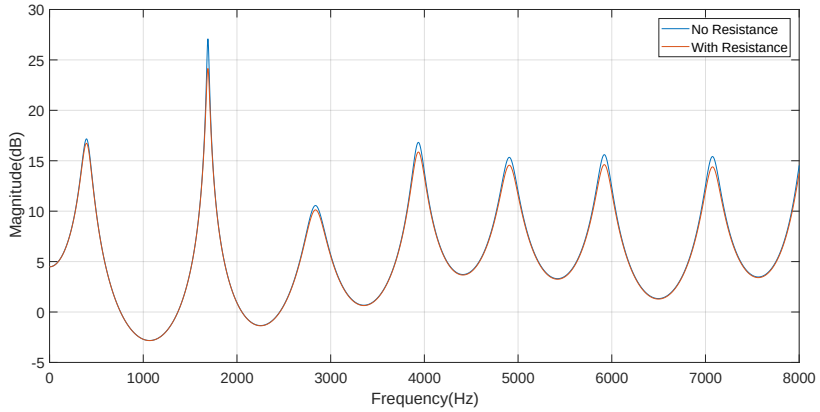


Figure 3.3: Effect of Resistance in T-network to Transfer Function

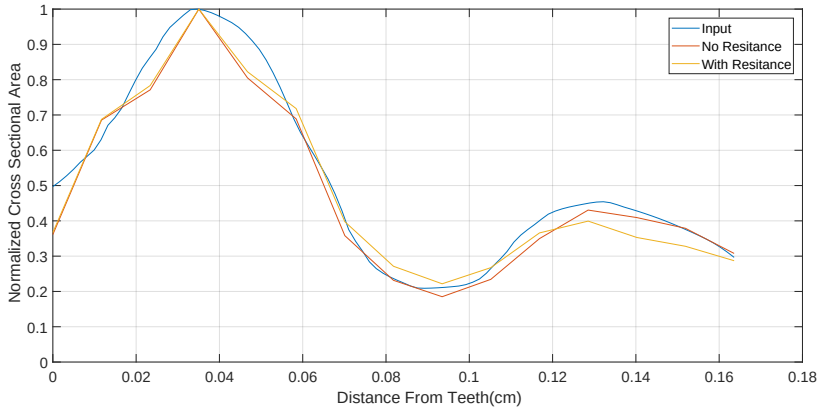


Figure 3.4: Effect of Resistance in T-network to Reconstructive Inverse Filtering

It can be seen that resistance decreases the amplitude of peak frequencies in the transfer function. Most of the spectrum is not impacted by the resistance and the formant frequencies stay the same. As a result, reconstruction is almost identical in cases with resistance and without resistance.

3.2.2. Wall Parameters

Wall impedance values are measured by Ishizaka, French and Flanagan [48]. Ishizaka et a1. made direct measurements of the vocal tract wall impedance of the cheek and the neck in vivo setting. The parameters that have been found can be seen in Table 3.1.

	M (kg/m^2)	B ($kg/m^2 s$)	K ($kg/m^2 s^2$)
Cheek, Relaxed	21	800	84.5×10^4
Cheek, Tense	15	1060	33.3×10^4
Neck	24	2320	491×10^4

Table 3.1: M,B,K wall parameters given in [48]

Additionally, Harper et al. [44] used measurements from studies of Suki et al.[76] on wall impedance of calf trachea in 1993 and Habib et al.[43] in 1994 on humans, in his model for sub-glottal impedance calculations. The list of the used parameters can be seen in Table 3.2.

	Tissue Density $\rho, (kg/m^3)$	Viscosity $\nu, (kg/ms)$	Elasticity $E, (g/ms^2)$
Value	1060	160	39200

Table 3.2: Default wall parameters given in [76],[44] and [43]

In Figure 3.5, the transfer functions of the vocal tract with reflection coefficient of glottis, $\Gamma_{glottis} = 0.7$ are shown for the different additional parameters as presented in Tables 3.1 and 3.1 with different formulations mention in Section 2.2.2.

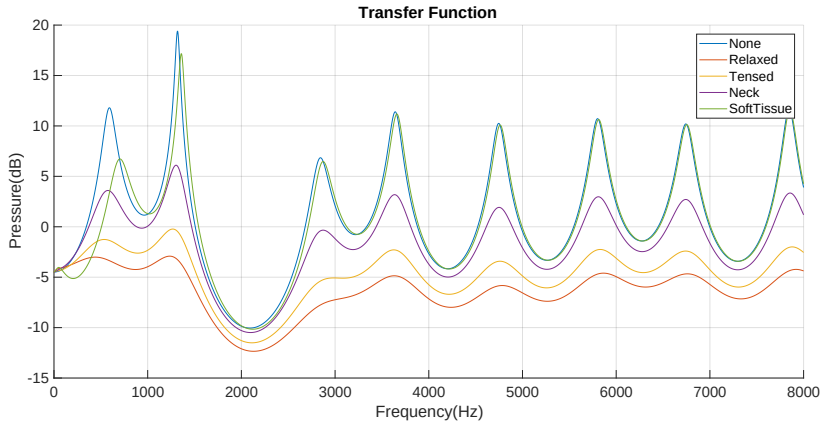


Figure 3.5: Effect of Wall Parameters in T-network to Transfer Function

In Figure 3.6 we have the results from the reconstruction procedure explained in 3.1.1 with the same aforementioned settings for model order. We see large differences in the reconstructed cross sectional areas.

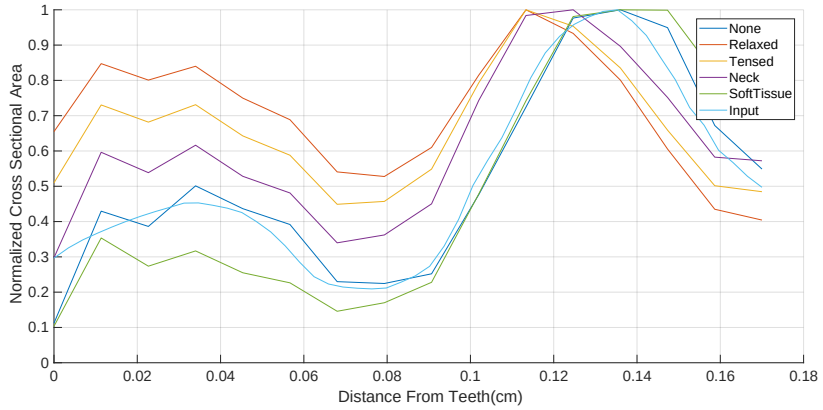


Figure 3.6: Effect of Wall Parameters in T-network to Reconstructive Inverse Filtering

As a result, it can be said that the model that can fully reconstruct the airway tract with linear prediction is the one suggested in Harper et al.[44][45]. The other models significantly alter the transfer function and in turn disrupt the reconstruction. Hence, for our models, we rather chose to select the model with wall parameters in Table 3.2 on the preceding page as the research is recent[76], and the methods are experimentally validated in various other studies [45] [44] [76].

3.2.3. Conductance

Conductance, caused by the the heat conduction within a duct can be modeled as a shunt resistor as defined in Section 2.2 on page 18. In Figure 3.7 we have plotted the vocal tract transfer function spectrum with and without conductance element. Tube is assumed to be open ended and reflection coefficient of glottis is taken as 0.7.

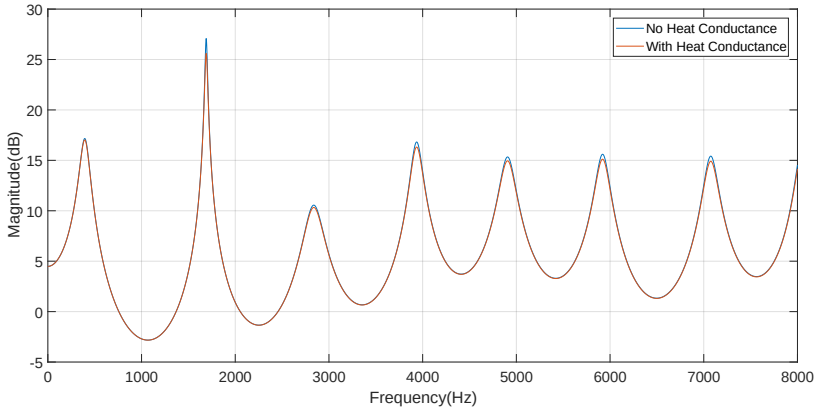


Figure 3.7: Effect of Conductance in T-network to Transfer Function

The reconstruction of the cross-sectional areas has been done using LPC with settings defined in section 3.1.1. We see the results in Figure 3.8.

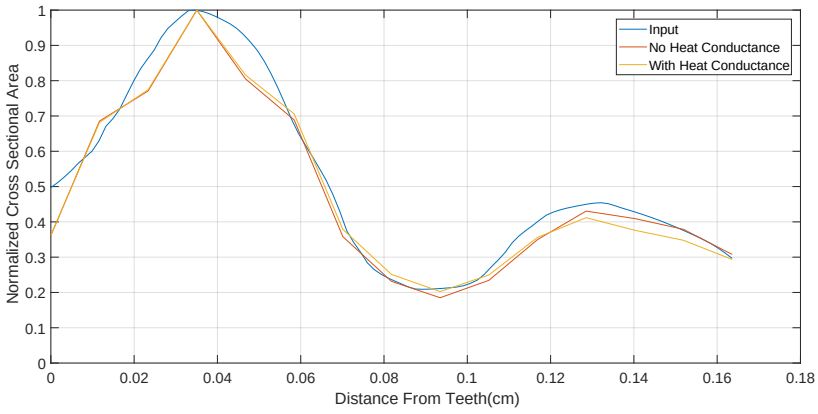


Figure 3.8: Effect of Conductance in T-network to Reconstructive Inverse Filtering

The results show that conductance has a small impact on the transfer function and the reconstruction is not significantly altered.

To conclude the matching of the tube model with linear prediction we state that under ideal conditions mentioned in 2.1.4 linear prediction can be used to estimate change in cross-sectional areas. However, these conditions are not ideal for tube airway for resistance, conductance and wall parameters. In this section we have seen that despite non-ideal conditions reconstruction is accurate concerning the formant frequencies, bandwidth, and error in reconstructed cross-sectional area.

3.3. Obstruction and Transfer Function

The main objective of our study is to find the impacts of obstruction on the transfer function and to establish if certain characteristics in the spectrum can be used to classify or detect an obstruction in the airway. To that end, we will impose different types of obstruction in the airway. This section will demonstrate how the transfer function is impacted by changes in cross-sectional area and changes of location of excitation. It shall be noted that location of obstruction and the location of excitation might be different sites, as they have different definitions [69]. These sites may or may not coincide in individual patients. In this study, we exclusively focus on the impact caused by the site of excitation. It is assumed for this study

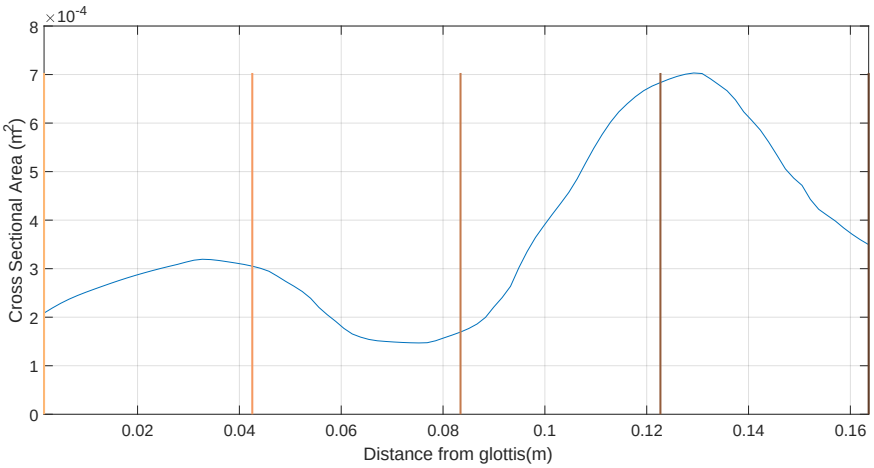
that, excitation is caused by the vibration of the narrowest regions in the airway.

In order to impose different obstructions as smoothly as possible, the obstruction area, surrounding obstruction location, has been divided to 4 equally distributed sections. The segment boundaries are taken as the points where the tube is constricted. Rest of the points within the obstruction area are interpolated. The resulting interpolated points are chosen as narrowed cross-sectional area function. This can be better understood in Figure 3.10a on page 35.

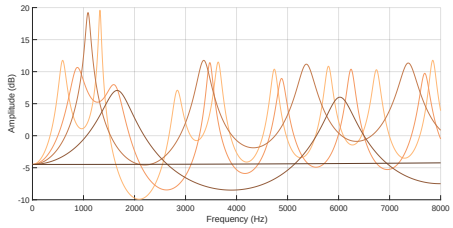
3.3.1. Location Of Obstruction

In this section we change the obstruction location from glottis to upper parts of the airway tract without changing the cross-sectional area profile. The experiment results can be seen in Figure 3.9 on the following page. The vertical lines represent the excitation location applied to the specified model. The response to the excitation location is color coded according to excitation location.

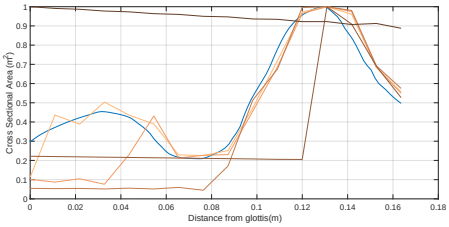
The model that has been taken into account in the following figures are the tube models mentioned in Section 2.3 on page 20. Lip Radiation has not been taken into account in order to observe the effect of linear prediction performance on the spectra created by the tube models. Reflection coefficient has been taken as 0.7 and additional parameters have been ignored to solely observe the singular effect of excitation location.



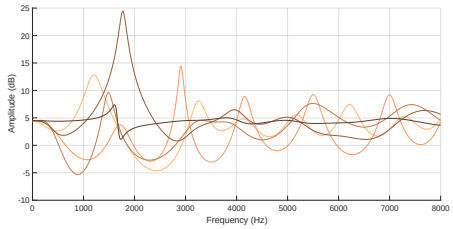
(a) Excitation Location(vertical line) Changing from glottis to Mouth



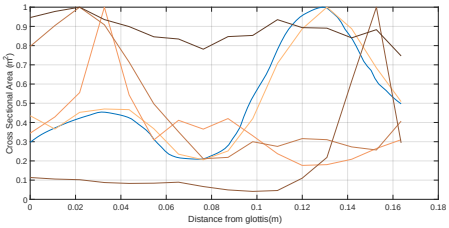
(b) Transfer Functions for Changing Excitation Location(vertical line) for Model 1



(c) Cross-Sectional Area Reconstruction for Changing Excitation Location(vertical line) for Model 1



(d) Transfer Functions for Changing Excitation Location(vertical line) for Model 2



(e) Cross-Sectional Area Reconstruction for Changing Excitation Location(vertical line) for Model 2

Figure 3.9: Impact of change of excitation location(top) for Model 1(Second Row) and for Model 2(Third Row)

It is also complementary to mention that, moving the excitation location to the ending of the tube results in a higher fundamental frequency for Model 1. This can be followed from Equation 2.31 on page 12 as decreasing the tube length results in higher formant frequencies. It can be seen in Figure 3.9c that for Model 1, reflection coefficients are close to 0 for locations behind the excitation location due to the way the model has been built. And in return, reconstructed airway is not changing behind the excitation location. It is evidently questionable to make this modelling operation for noisy measurements, as spectrum estimation is not as detailed for ideal tube modelling we present here.

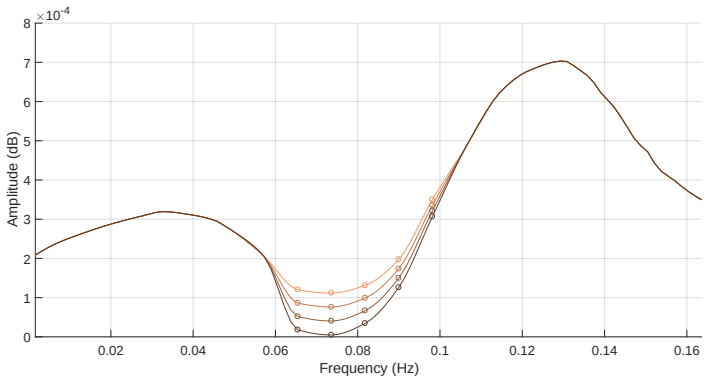
As a result, it is possible to estimate the excitation location by observing the reconstructed cross-sectional area, or reflection coefficients. However, for the second model, as impedance behind the excitation location is taken into account, we can not reconstruct the cross-sectional area of the upper airway. This result shows clearly that inverse filtering can be plausible if and only if we assume that the obstructing source can be taken as an anechoic ending.

It should also be mentioned that since we do not know the excitation location, we have

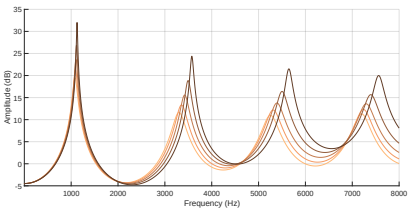
determined the order of the linear prediction using the order defined in Equation 3.1 on page 28. The selection of the order is a causality dilemma, as order can not be determined for unknown obstruction location. Hence, if the model 2 is closer to reality, and is the determining model for transfer function for the upper airway tract during snoring, determining cross-sectional areas from LP is not an option.

3.3.2. Obstruction Severity

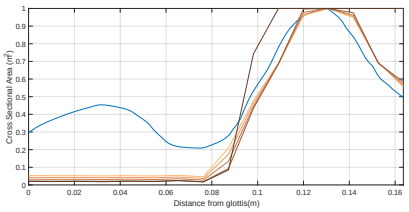
The severity of the obstruction can impact the transfer function, and in turn, the reconstruction of cross-sectional areas. In this subsection, we increase the obstruction severity of the airway tract and observe the spectrum change with increased intensity. We have chosen the narrowest part of the airway tract and decreased the cross-sectional are for the surrounding anatomical region. The results for Model 1 and Model 2 can be seen in Figure 3.10.
The inverse modelling has been done using the same order in section 3.1.1. Again, we come across causality dilemma mentioned in the previous section.



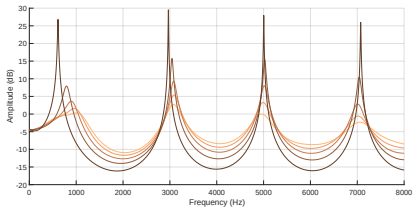
(a) Varying Cross-Sectional Areas for different obstruction severity



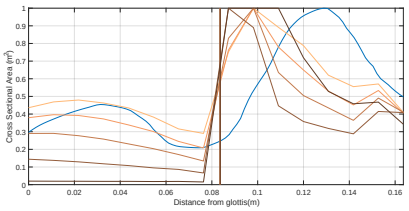
(b) Transfer Functions for Changing Obstruction Severity(vertical line) for Model 1



(c) Cross-Sectional Area Reconstruction for Changing Obstruction Severity(vertical line) for Model 1



(d) Transfer Functions for Changing Excitation Location(vertical line) for Model 2



(e) Cross-Sectional Area Reconstruction for Changing Obstruction Severity for Model 2

Figure 3.10: Impact of change of Obstruction Severity(top) for Model 1(Second Row) and for Model 2(Third Row)

Results in Figure 3.10 can be expected from equation 2.54 on page 15. As it can be seen from T-matrix, the frequencies are mainly impacted by the length of the tube, l. However, magnitude of the peaks are impacted by the cross-sectional area of the tube, S. As a result, main impact of obstruction severity is increase of magnitude of resonant frequencies. This

points that it is not possible to measure the degree of severity of the obstruction from spectral characteristics, as the experiments are not precise due to nature of sound recording. A slight decrease in the resonant frequencies is observable for Model 2.

Consequently, LP is not an adequate way to detect the severity of the obstruction even if we assume that Model 1 is the model that reflects the physical reality.

3.3.3. Velum, Oropharynx, Tongue, Epiglottis Snoring Simulation

Here, we will introduce the configurations mentioned in Kezirian et al. [53] applied to cross-sectional area findings obtained from Mohsenin et al. [60]. While the definition of Kezirian et al. for regions that may include constriction depends on movement of these tissues in three dimensions, our configuration is bounded by movement in two dimensions.

We in turn created four different profiles for different regions mentioned in section 1.1 in the study of Kezirian et al. [53]. From Figure 3.11, it is evident that even though different excitation locations are effective at snoring sound production, they are located very close to each other within the upper airway.

In Figure 3.12 we see these profiles that correspond to Epiglottis, Velum, Tongue, Oropharyngeal Walls color coded from light to dark. The spectra of the transfer functions can be seen in Figures 3.13 3.14.

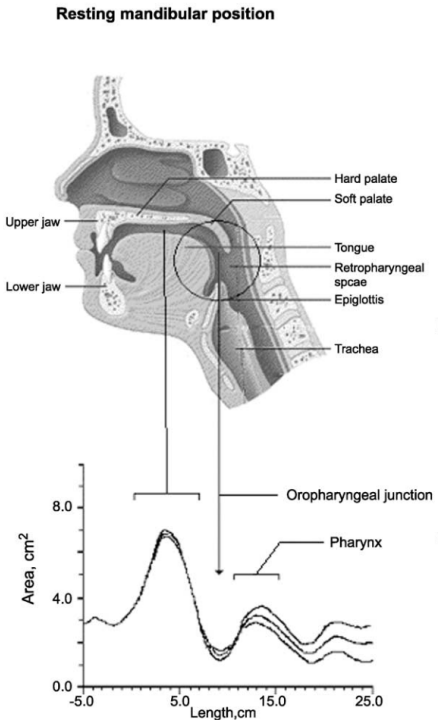


Figure 3.11: Upper airway cross sectional area during sleep using acoustic reflectometry, from Mohsenin et al.[60]

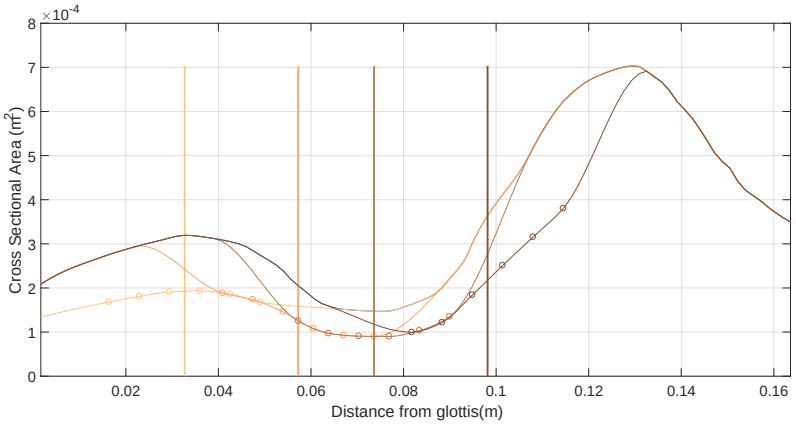


Figure 3.12: Cross-Sectional Area for VOTE classification. It is assumed that the effected structures are (from left to right vertical lines) Epiglottis, Velum, Tongue, Oropharyngeal Walls

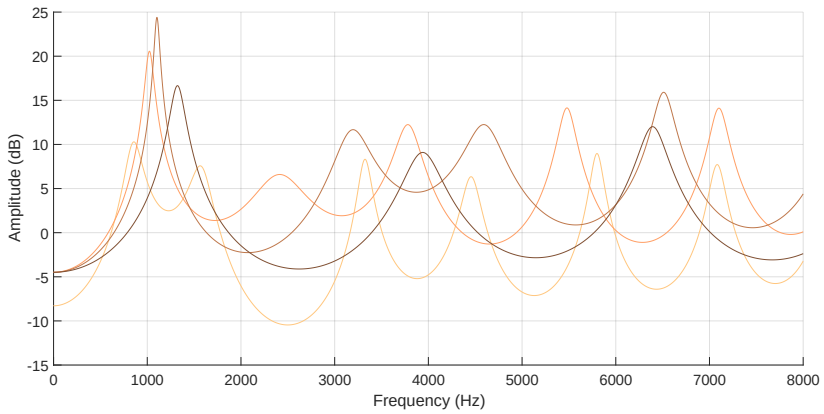


Figure 3.13: Transfer Function of Tube Model 1

It can be seen in Figure 3.14 that spectra for model
2 has fundamental frequencies below 500 Hz. This is in accordance with the research on acoustic properties of snoring [29] [70] [12] [1] [10].

3.4. Synthetic Speech

We first aim to synthesize clear speech sounds in order to validate our model and software. Later, we will use the same framework to synthesize snoring sounds.We are going to analyze speech using inverse filtering on real life pronounced speech signals from TIMIT Acoustic-Phonetic Continuous Speech Corpus [39] in section 3.4.1. From phonetic list in TIMIT, we have selected the vowels that were used in [78] which are shown in Table 3.3.

Table 3.3: Vowels inspected in TIMIT with their ARPABET correspondence

2 Letter ARPABET	International Phonetic Alphabet	Examples
IY	i	beat
EH	ɛ	bet
AA	ɑ	balm, bot
AO	ɔ	bought
UW	u	boot

From the inverse filtering method, we will extract the vocal tract transfer function and

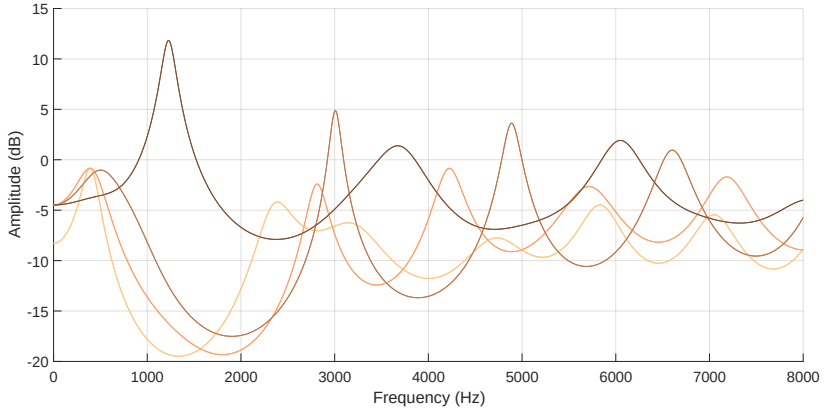


Figure 3.14: Transfer Function of Tube Model 2 for excitation locations in Figure 3.12 on the facing page

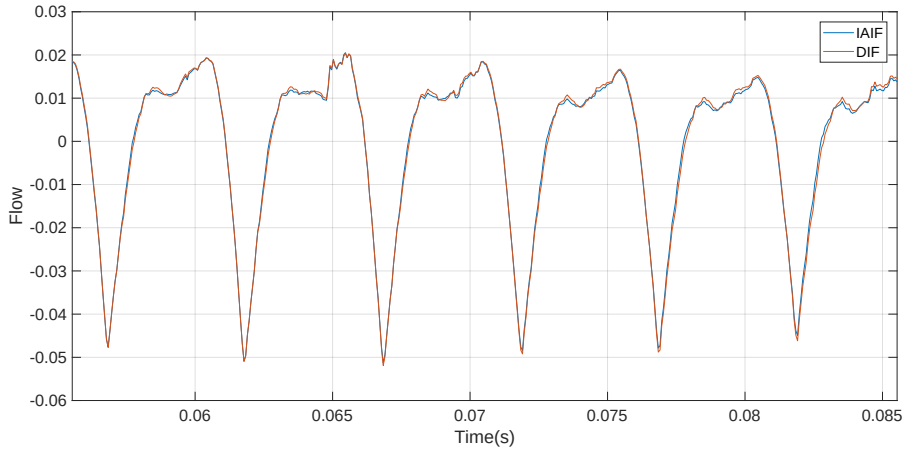


Figure 3.15: Inverse Filtering Model Comparison for ɑ

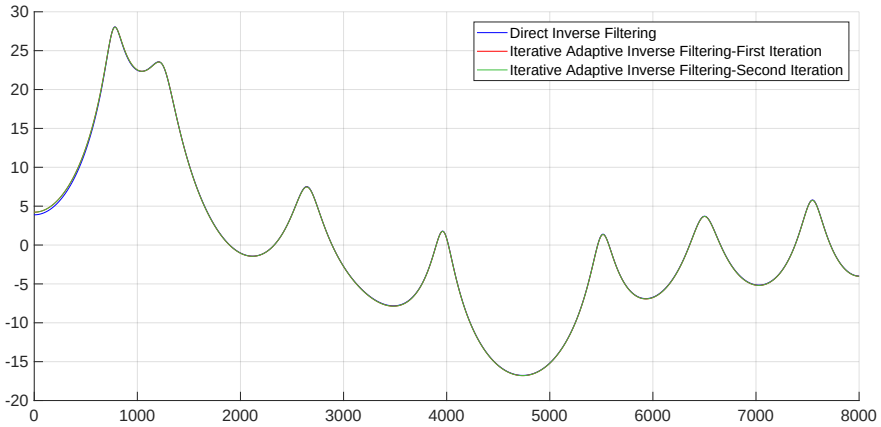


Figure 3.16: Comparison of two inverse filtering methods and respective transfer functions of the vocal tract

compare it with the tube model transfer function in section 3.4.2.

3.4.1. Inverse Filtering

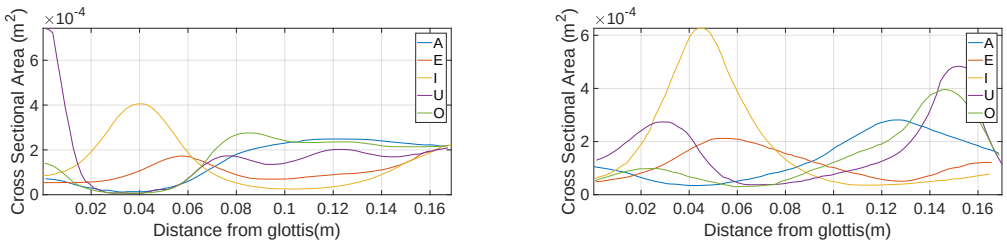
As mentioned in section 2.5 we applied two different inverse filtering methods to find the excitation flow waveform for snoring. First, sound files with orthography "She had your dark suit in greasy wash water all year." has been selected to extract phonemes introduced in Table 3.3 on the preceding page. Then, for each vowel, both of the inverse filtering techniques mentioned in section 2.5 on page 24 has been executed. The results for a sound excerpt of ɑ can be seen in Figure 3.15. Comparison for the other vowels can be found in Appendix A

Form Figure 3.15, it can be said that waveform for IAIF and DIF are almost identical. As a result, transfer functions for the vocal tract are also identical as seen in Figure 3.16, with a little difference in lower frequency region. Hence, we will rely on the simpler method, i.e. direct inverse filtering for our further calculations.

3.4.2. Vocal Tract Transfer Function

Here, we will discuss how we obtained the vocal tract transfer function from the cross-sectional areas. This section does not include lip radiation or the effect of the glottis. The cross sectional areas of the airway are obtained from Mermelstein et al. [58], which include Russian and American vowels and Wakita et al. [78]. As a result, for Model 1, we expect to find

the same results in aforementioned studies. Cross sectional areas and the resulting transfer function spectrum of vocal tract for vowel ɑ has been given in Figure 3.18. The cross sectional areas are different for multiple reasons: Study of Fant [34] is conducted with X-ray techniques to directly determine the human vocal tract shape. Mermelstein found that using first six formant frequencies is sufficient to determine the shape of the vocal tract using asymmetric approximation. Wakita's cross sectional functions are derived from sound files using lattice filter LP model, and it has been discussed that these are found to be 'reasonable' compared to Fant's results.



(a) American Vowels from Mermelstein et al.[58] (b) Russian Vowels from Fant [34]

As mentioned in Section 2.3 on page 20, there are two models that can be selected to model the vocal tract. In this section, as our objective is to confirm the validity of the created software and the previous studies on the vocal tract and speech production, we used Model 1 to create the transfer function spectrum, without including the lip radiation. Transfer functions for other vowels can be found in Appendix C on page 69. For all the vowels, the transfer functions do match the aforementioned studies.

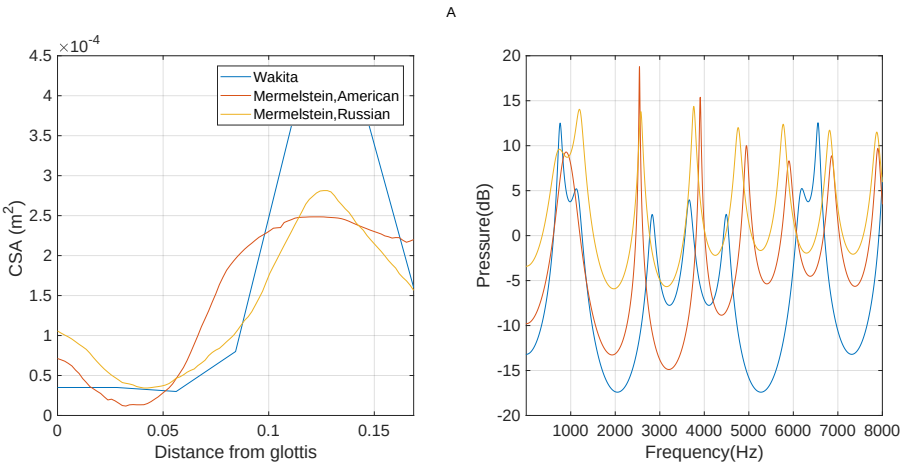


Figure 3.18: (left) Cross-sectional Areas; (right) Transfer Function for vowel ɑ

As the next step, we will compare the results from acoustic transfer function with transfer function obtained from the linear prediction coefficients estimated in the inverse filter. We chose three different segments within a vowel; start of the vowel sound (the first frame), center of the vowel sound (middle frame), and end of the vowel sound(last frame). Cross-sectional areas for Russian vowels in [58] has been discarded in figures below, as the TIMIT database constitutes of American vowels. The figures below are taken from a set of vowel segments and averaged per frequency.

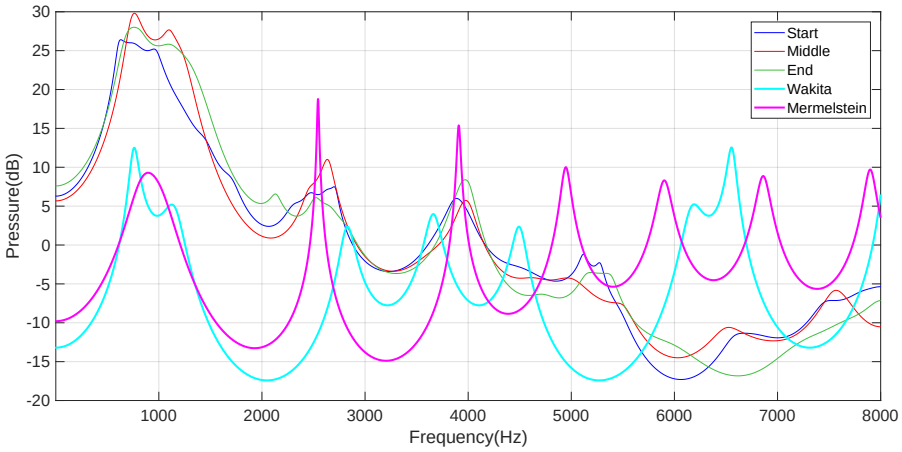


Figure 3.19: Comparison of Vocal Tract Transfer functions from Inverse Filter vs Tube Model for vowel ɑ

It can be seen that there exists rough correspondence between start, middle and end frames for a vowel sound. The higher frequencies are less congruent. Correspondence between Wakita and Mermelstein on American vowel sounds is significant for lower frequency range, $f < 3000$. Higher formant frequencies are located differently. Bandwidth of the formant frequencies are also similar. However, we see that despite cancelling the combined effect of the glottal flow and the lip radiation through inverse filtering we see that there is still a decrease in amplitude for increasing frequency.

Here, it shall also be noted that vocal tract length of the specific subject has a high impact on the shape of the spectrum. Essentially, longer vocal tract results in lower resonant frequencies according to Equation 2.31 on page 12. Additionally, dialect can impact the shape of the spectrum due to different intonation of the vowels. As mentioned in Story et al.[9] these voice qualities change the vocal tract shape. It can be seen in Figure 3.19 that formant frequency in the lower frequency range $f < 5000Hz$ do match for tube model and inverse filter. However, precision in higher frequencies is lower.

3.4.3. Pitch Estimation

In order to find an average waveform, pitch of the sound signal has to be estimated. YIN pitch estimator [25], has been selected due to it simplicity, no parameter requirement, and low latency, in addition to resilience to low SNR[25]. It forms the scalar product between two windows that shift symmetrically in time with respect to the analysis point and uses autocorrelation method and tweaks it with modifications to avoid estimation errors. The window size was 40 ms, the threshold parameter was 0.1, and the F0 search range was 25 Hz to one quarter the sampling rate which is 4 kHz for our database. The window shift was same with window size, i.e. 40 ms.

3.4.4. Peak Finding

A simple peak finding algorithm from Imperial College Voicebox [21] has been used in order to sync the peaks for a vowel sound between different pitch periods. These peaks are used to extract pitch periods for a certain vowel. Later, we take the median for each sample to find a waveform estimate for a specific vowel.

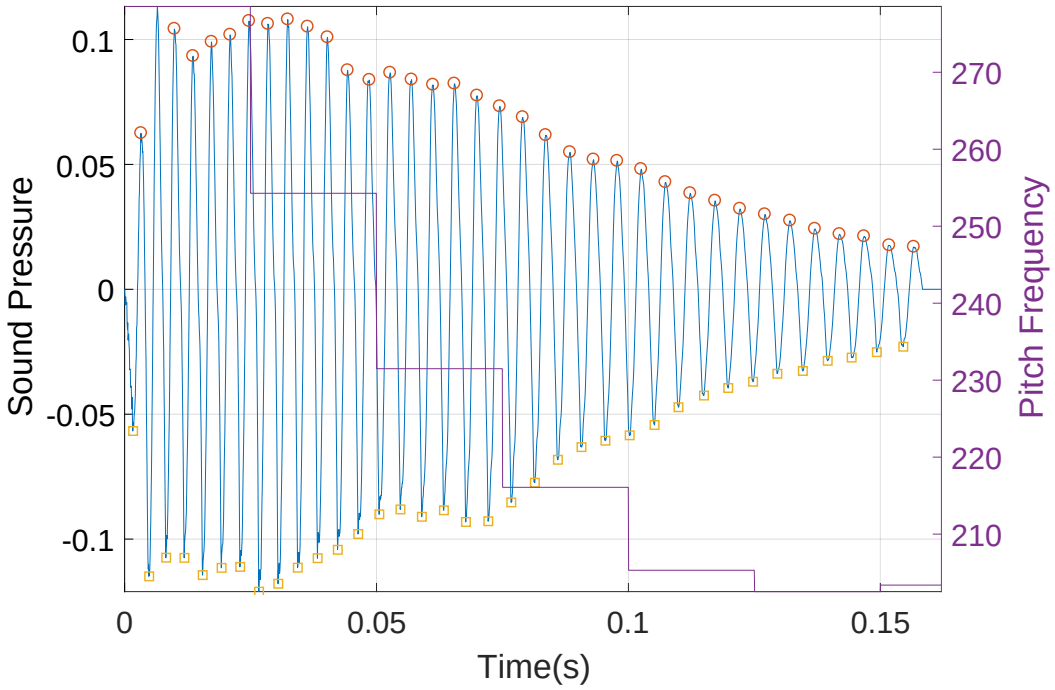


Figure 3.20: Peaks and Pitch Period for a glottal flow signal analyzed from ɑ

In figure 3.20 we can see that peak finding algorithm works well in speech signals as the peaks and the valley are marked with red and yellow circles respectively. Additionally, extracted pitch frequency reflects the well-known declining trend of F_0 in speech [25]. As a result, the glottal waveform and the glottal derivatives can be determined with relatively small confidence intervals 3.21, 3.22 on the next page.

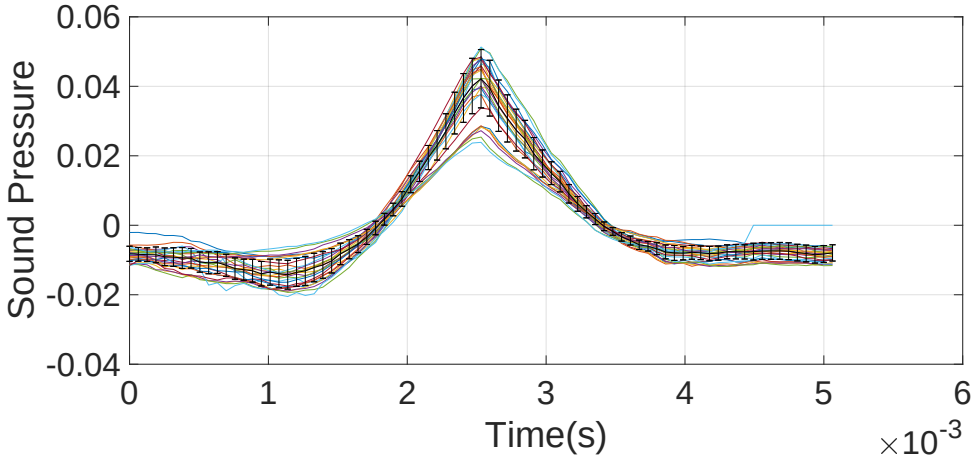


Figure 3.21: Glottal Waveform for ɑ

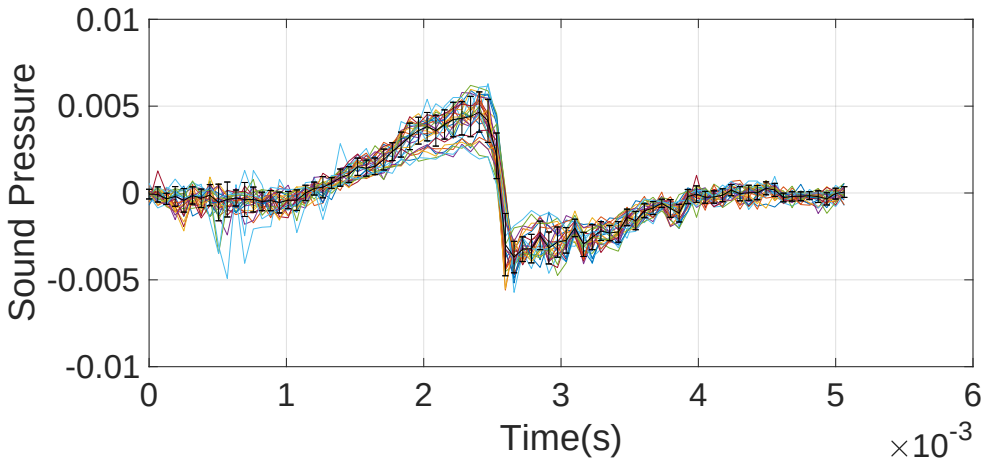


Figure 3.22: Glottal Derivative Waveform for α

As shown in Figure 3.21 on the preceding page, 3.22 we obtain a robust result for glottal flow and glottal flow derivative. The error spread is low and it matches the previous literature on glottal signals[61].

3.4.5. Lip Radiation

As mentioned in section 2.1.6 on page 16 lip radiation has an impact in the higher frequencies of the spectrum. We can compare the impact of differentiation and complex models mentioned in the Lip Radiation section as seen in Figure 3.23. The model that has been chosen for lip radiation for speech synthesis, has a minimal differential impact compared to a differentiation besides gain difference. Here, we used glottal flow derivative in order to disregard the lip radiation effect, as lip radiation is commonly modeled as derivation in Source Filter Model.

3.4.6. Overlap-Add

In order to synthesize a speech signal, after finding an average flow waveform for an average pitch period, we used a simple overlap-add method to synthesize a longer, audible vowel sound. Pseudocode for overlap add is given in Algorithm 2 on the next page.

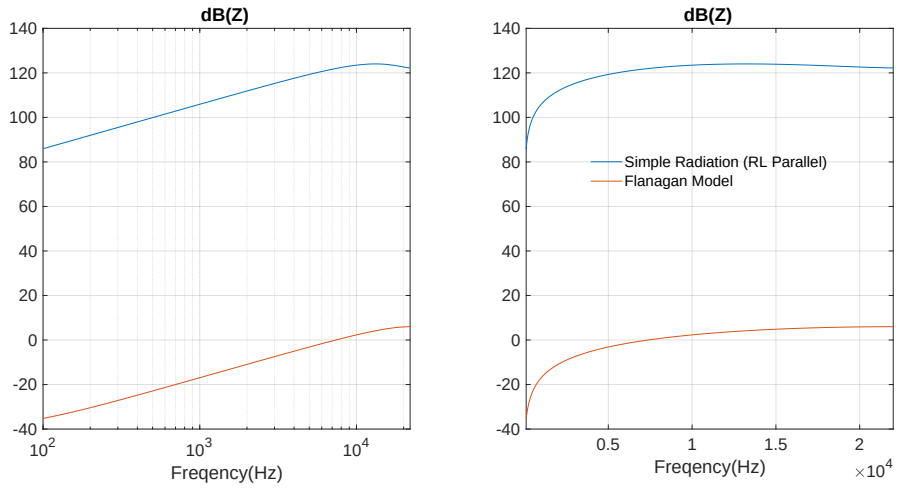


Figure 3.23: Lip Radiation with Derivation and Struve Function approximation [1]

input : Single waveform x,length of window $winLen$,
shift,i.e. pitch frequency, $winShift$
output: Overlap added signal y

```
win = hanning(winLen);
while stop <= length(g) do
    // Overlap and add
    y(start:stop) = y(start:stop) + x * win;
    // Increment
    start = start + winLen - 1;
    stop = start + winShift - 1;
end
```

Algorithm 2: Overlap-Add with Windowing

3.4.7. Synthetic Speech Evaluation

Speech signals have been synthesized from the average glottal flow derivative signals and convolved with impulse response of the transfer functions of the vocal tracts in Figure 3.18 on page 39 and ones presented in Appendix. The results from the previous sections has been evaluated by speech expert in Koninklijke Philips N.V. Research Department. The results were found to be satisfactory considering the following factors in speech production method:

1. Pitch Jitter: The signal has been synthesized with constant pitch period as an ensemble average for multiple pitch period has been calculated. As the real flow signal is quasi-periodic, synthesized signal does not sound as natural as the real life signal. This can be compensated by adding randomness to window shift in overlap-add methods. However, as there hasn't been a study for pitch jitter distribution, this modification has been disregarded for our study.

2. Amplitude Shimmer: Amplitude of the flow signal may vary for different pitch periods. As we do overlap-add synthesis for an averaged signal, the amplitude shimmer is not reflected in the synthesized result. Amplitude Shimmer can also be compensated by multiplying the windowed waveform with a random gain. However, as there hasn't been a study for this probability distribution, this has been disregarded.

3. Aspirated voicing, turbulence might occur due to nonlinearities in the airway tract, or an opening in the closed phase of the glottal waveform. This can be modeled by adding random noise to excitation waveform.

3.5. Synthetic Snore

In order to evaluate inverse filtering and tube modelling on snoring sounds, we will use real-life signals to analyze excitation flow of snoring signals. In section 3.5.1, the data that have been used in this analysis will be introduced. Then, the framework for the snoring sound synthesis will be mentioned. We will also investigate a simple Voice Activity Detector and evaluate success of YIN pitch detection on snoring sounds in section 3.5.2. Later, in section 3.5.3 on page 47, the shape of the average waveform for snore will be evaluated and compared with glottal waveform.

3.5.1. Crossover Trial

In this section, we will experiment on snoring sounds that have been collected in Silent-Night project in United States, Philips Respironics. The trial was conducted in order to evaluate effectiveness of different anti-snoring devices and to find a method to determine the best solution for the couple. A preselection has been made to exclude patients with OSA. 27 participants, comprising of a snorer, and if available a snoring partner, used three different

devices for 38 days. The device effectiveness was assessed subjectively after each night and after every device period. The sound recordings were made with a mobile phone (Motorola G5) with a java app, and encoded in OGG Vorbis format. The timeline for the trial is given in 3.24. For MAD and PD, a period of one week is dedicated for adjustment and training. Participants were interviewed after each device usage, specifically during day 10,24 and 38.

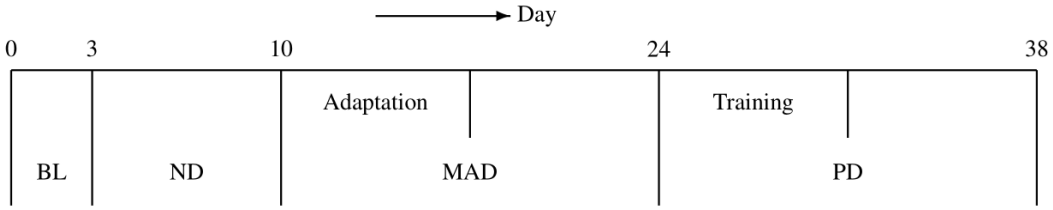


Figure 3.24: Timeline of Crossover Trial, Baseline Period (BL), Nasal Dilator (ND),Mandibular Advancement Device (MAD),Positional Device (PD)

The collected data consisted of information at intake and exit, daily surveys, the interviews at crossovers, and the audio recordings. The intake, exit and crossover interviews were extensive: the number of questions for snorer and partner were (25,26), (17,17) and (17,19) for the intake, crossover and exit interview, respectively. The daily surveys comprised of 8 questions, including a score of the sleep quality rated by the *snorer*, and scores by the *partner* for disturbance, snoring loudness, and sleep quality.

The snoring periods are detected by a dedicated algorithm developed in Philips Research by Brinker et al. [18] in Java Environment. The algorithm uses autocorrelation to find breathing rate and finds the volume difference by k-means clustering. Later, several conditions are applied to detect snore epochs. Two of the important conditions are minimum and maximum breathing rate (4 BPM,20 BPM respectively), 4dB threshold for threshold level between K-means segments. Segments are 0.5 seconds long (8192) and shifted by 0.125 seconds (2048). The program outputs in warped linear prediction features, binary snore classification, k-means relative energy, breathing rate with 8Hz sampling rate. Additionally, program outputs snoring epochs as Waveform Audio File Format with 16000 Hz. These *snoring epochs* are a period of consecutive *snore events* and used in our snore analysis.

As a result of the trial, patients who benefit from a particular device have been identified in Brinker et al. [19]. In this study multiple sources of evidence for the device effectiveness has been considered. As the first source of supporting evidence we considered the daily subjective data on loudness of snoring, sleep disturbance and sleep quality of the partner, which are daily inputs. As the second source of supporting evidence we considered the daily objective data on loudness and duration of snoring as measured by the aforementioned Java app. As the third source of supporting evidence we considered the same daily objective data on loudness and duration of snoring but now did a comparative analysis between device periods. For all of these criteria, Welch Test [81] have been used to identify statistical significance.

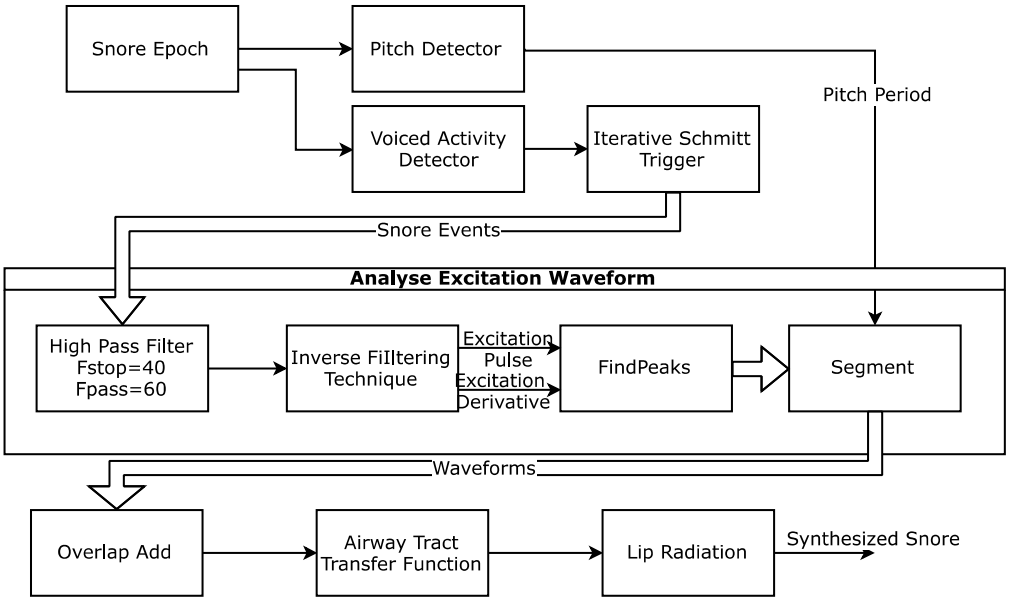


Figure 3.25: Excitation Flow Analysis Framework

Our framework follows the procedure is explained below.

1. A snore epoch from a specific patient on a specific night is taken. These sound excerpt can be between 4 minutes to 2 hours.

2. We detect the pitch with YIN pitch detector for the entire file.

3. We use a voice activity detector using spectral flatness of the signal 3.5.2. The threshold for spectral flatness is taken to be 0.3;

4. An iterative schmitt trigger is used to embed hysteresis effect. Output of schmitt trigger is multiple 1-2 second snore events that occur during inhale periods.

5. A High pass filter is used to eliminate low-frequency fluctuations captured by the microphone

6. An inverse filter mentioned in section 2.5 on page 24 is used to obtain excitation flow for a snore event

7. The peaks are found using the algorithm mentioned in section 3.4.4 on page 40.

8. Segments of pitch period duration are taken, centering the maximum amplitude of excitation waveform. We later interpolate the events with different length and find an average waveform for the snore epoch.

9. Overlapp-Add with a fade in and fade out window is applied to synthesize an excitation flow signal of longer duration.

10. The signal is convolved with impulse response of the airway tract transfer function

11. If the excitation pulse is taken to be the main signal, signal is convolved with impulse response of the lip radiation. If excitation derivative is taken, this box is ignored.

3.5.2. Voiced Activity Detector and YIN Pitch Detection

Here, we used a voiced activity detector that incorporates spectral flatness. In theory, spectrally flat frames are prone to be silent, as they would have noise characteristics. We calculate Welch's power spectral density estimate. Welch's method computes a modified periodogram for each segment of length 0.025 seconds and then averages these estimates to produce the estimate of the power spectral density. Later, we calculate the spectral flatness by finding the geometric mean and divide to mean of the spectrogram. Later, we normalize by dividing to maximum flatness. The threshold for a silent frame is that, normalized spectral flatness has to be larger than 0.3.

YIN Pitch detection algorithm is applied over the whole signal. The results are seen in Figure 3.26 and 3.27. Most of the detected segments are identical using the Exponential Smoothing and Schmitt Trigger smoothing. It has been tested that Schmitt Trigger is faster than exponential smoothing. Hence, Schmitt Trigger is used for computational efficiency.

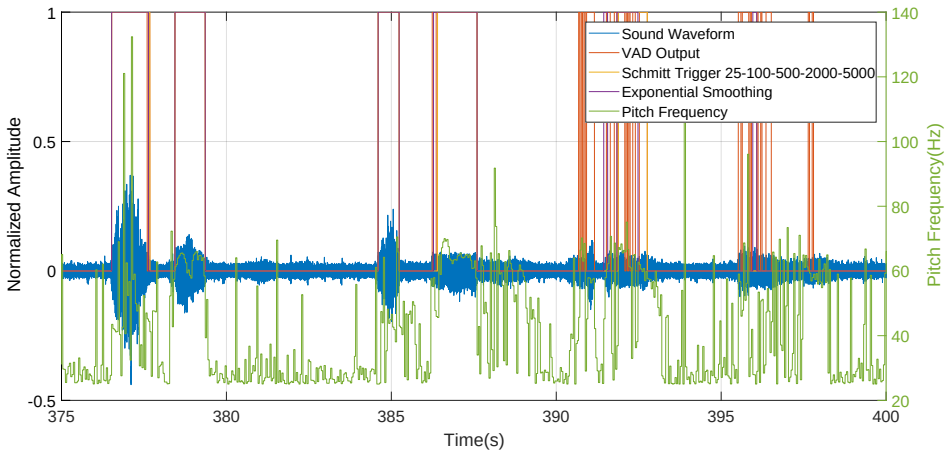


Figure 3.26: Histogram for Median and Mean Pitch Frequencies for a Snore Epoch

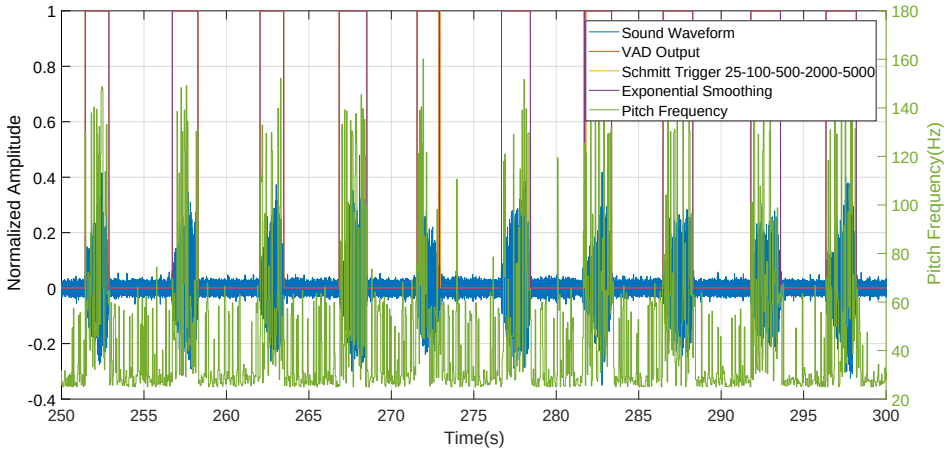


Figure 3.27: Histogram for all voiced Pitch Frequencies for a Snore Epoch

It is required to mention that a snoring sound during inhalation is usually 1-2 seconds long, while a vowel sound is around 20 milliseconds long. Hence, it is beneficial to state that as mentioned in Huang et al. [47], the amplitude of the snoring signals increases as the amount of pressure increases with time. Hence, pitch detection algorithm detects different

pitch for different time frames within a signal. Here, it has to be mentioned that the snoring sounds that have been used in this trial have a flat spectrum due to inclusion of white noise caused by both imperfect mobile microphones and the distance between the audio source, the snorer's mouth, and the microphone. This causes the pitch to be erratically estimated even though YIN pitch detector is resilient to low SNR. It can be seen in the histograms below and that despite it might be unstable within a single snore signal (see: 3.26 on the preceding page), if we take the median for each snore event, we can find an estimate around the range 60-80Hz. as seen in Figure 3.28

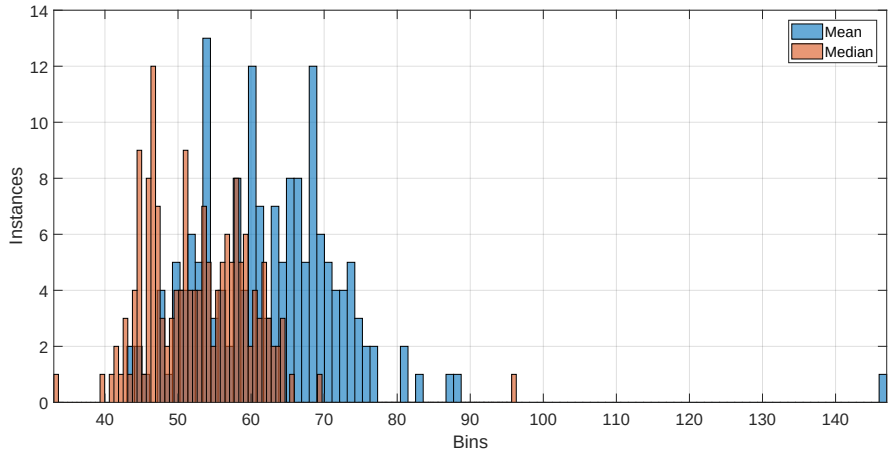


Figure 3.28: Histogram for Median and Mean Pitch Frequencies for a Snore Epoch

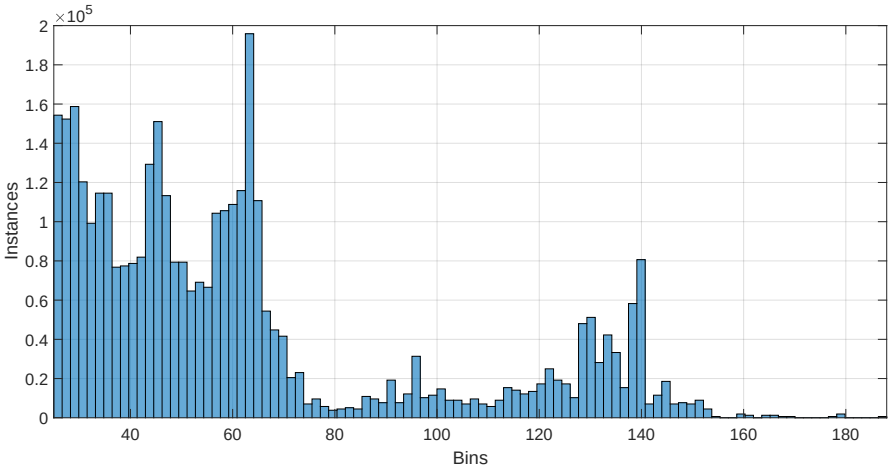
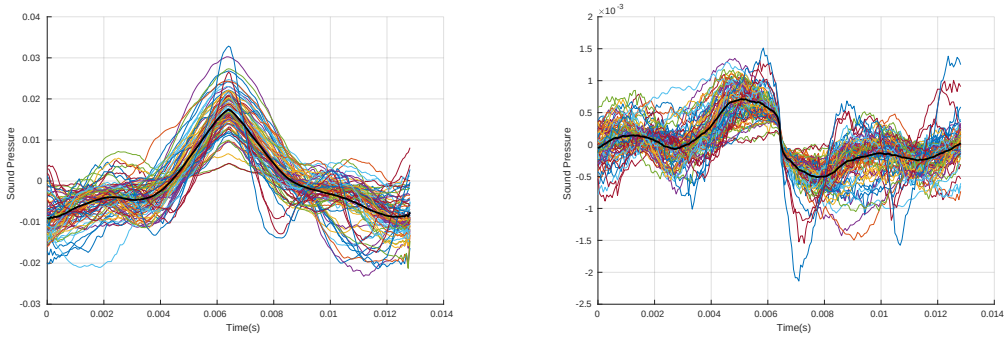


Figure 3.29: Histogram for all voiced Pitch Frequencies for a Snore Epoch

3.5.3. Average Excitation Waveform

As we have found the audible snore events and used filtering to estimate the excitation flow signal, it is now necessary to envelop the signals surrounding the peak amplitude. The results can be found in Figures 3.30a on the next page and 3.30b on the following page



(a) Excitation Waveforms for a Snore Epoch

(b) Excitation Derivative Waveforms for a Snore Epoch

Figure 3.30: Average Excitation Signal and Excitation Derivative Signal. Each line indicates a snore waveform with a pitch period. Black line indicates averaged waveform for a snore event

It can be seen in Figures 3.30a and 3.30b that the shape of the excitation flow is different for snoring and speech (Figures 3.21 on page 41 and 3.22 on page 42). The pitch of both vowel sounds is higher than the snoring signals. As it has been discussed in Section 3.5.2 on page 46 the pitch within a snoring signal is less stable causing the shape of the excitation signals to change within each snore instance. These waveforms are hard to compare with previous literature as it has not been experimentally investigated. In vivo experimentation while snoring for non-OSA population in order to find Cross sectional area change with an imaging device has not been present while this thesis is being written. The results however can be compared to work of Huang et al. [47] where consecutive frames of high-speed video filming of a collapsing tube conveying airflow has been taken and plotted. Despite the results do resemble each other, the lower frequency sampling in the study (1000 frames/sec) and lack of correspondence to natural setting makes the comparison less sensible to derive conclusions. Lastly, Ng et al. [61] followed a similar method to find snore excitation flow. The sourceflow results from this study matches the flow found in our study. Parametrization for snore flow signals can be found in Ng et al.[63].

3.6. Comparison of Inverse Filtering Transfer Function and Snoring Models

In this section, vocal tract transfer functions obtained from the inverse filtering methods applied to real snoring sounds, will be compared to tube model transfer functions. As we do not have the ground truth for the vocal tract transfer function we will take the spectra found by the inverse filter to be the empirical evidence as these spectra has been deducted by real-life snoring sounds.

The power spectrum obtained from these transfer functions will be compared in order to measure spectral distortion. We will assume that power specturm obtained from the sound files are the true spectrum of the vocal tract i.e. $P(\omega) = |H(e^{j\omega})|^2$ and estimated power spectrum of the vocal tract $P(\omega) = |H(e^{j\omega})|^2$, obtained from the tube model $\hat{P}(\omega) = |\hat{H}(e^{j\omega})|^2$.

Itakura Saito distance has been given by [49] [71]

$$D_{IS}(P(\omega),\hat{P}(\omega)) = \frac{1}{2M}\sum_{\omega=-M}^{M}\left[\frac{P(\omega)}{\hat{P}(\omega)} - \log\frac{P(\omega)}{\hat{P}(\omega)} - 1\right] \tag{3.2}$$

where M is the size of the fourier transform. It shall be mentioned that D_{IS} is a non-symmetric measure, i.e. $D_{IS}(P(\omega),\hat{P}(\omega)) \neq D_{IS}(\hat{P}(\omega),P(\omega))$. A symmetric measure has been defined by Gray and Markel [41] and evaluated in [42]. The measure can be defined by

$$D_{COSH} = \frac{1}{2}\Big(D_{IS}(P(\omega),\hat{P}(\omega)) + D_{IS}(\hat{P}(\omega),P(\omega))\Big) \tag{3.3}$$

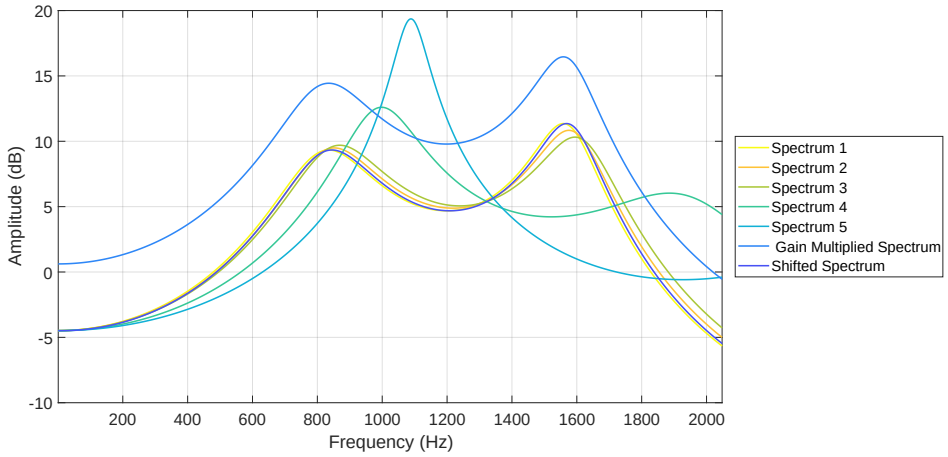


Figure 3.31: Examples of Itakura - Distance Measure

	1	2	3	4	5	GMS	Shifted
Spectrum 1	0	0 56	0.0228	0.3870	0.5594	0 00	0 16
Spectrum 2	0 57	0	0 60	0.3301	0.5227	0 57	0 17
Spectrum 3	0.0236	0 61	0	0.2744	0.4875	0.0236	0.0141
Spectrum 4	0.5180	0.4410	0.3634	0	0.2525	0.5180	0.4865
Spectrum 5	0.9286	0.8811	0.8319	0.3590	0	0.9286	0.9116
Gain Multiplied Spectrum	0 00	0 56	0.0228	0.3870	0.5594	0	0 16
Shifted Spectrum	0 16	0 18	0.0140	0.3705	0.5589	0 16	0

Table 3.4: Table for Distance Measures

We also introduce here a gain independent method to compare two spectra, Itakura Distortion measure. This is sensible as signal level makes no difference in the human understanding of sound.

$$D_I(P(\omega),\hat{P}(\omega)) = \log\frac{1}{2M}\sum_{\omega=-M}^{M}\left[\frac{P(\omega)}{\hat{P}(\omega)}\right] - \frac{1}{2M}\sum_{\omega=-M}^{M}\log\frac{P(\omega)}{\hat{P}(\omega)} \tag{3.4}$$

The gain difference between vocal tract transfer function from inverse filter and the transfer function from tube model is significant. As a response, for our experiments, Itakura Distance is an adequate measure, as it is gain independent[71].

In figure 3.31, we have created five different spectra; first three, very similar to each other, and fourth and fifth are different. We have also manipulated Spectrum 1 by multiplying the spectrum, and shifting the spectrum to the right. The results for the distance measures is given in Table 3.4

In Figure 3.31, it can be seen that gain has no effect on the distance measure, and the measure has low sensitivity to shifting the spectrum. As seen above, fifth spectrum is the most different to Spectrum 1 compared to other spectra, and this is reflected in the table above.

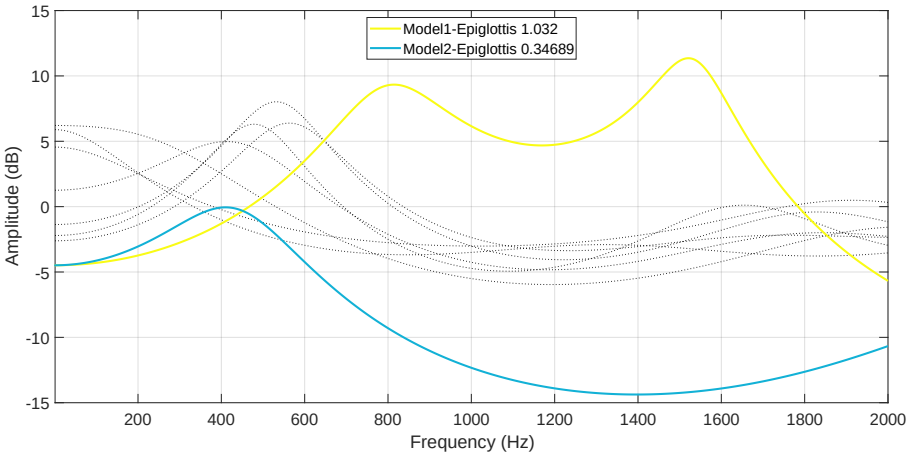


Figure 3.32: Spectra for tube models and spectra from inverse filtering. The numbers in the legend are corresponding averaged Itakura distances to randomly selected spectra of resulting inverse filter vocal tract transfer functions represented in black dotted lines

Spectral comparison and results for both methods are shown in Figure 3.32. For inverse filter method, a model order of 16 has been used, in concordance with section 3.1.1 equation 3.1 on page 28. Inverse filter method finds dominant peaks below 1000 Hz and in many of the cases, find below 500 Hz. However, it can not be replicated by a tube model 1 as looking at Equation 2.31 would show that such a tube would require double the length of average airway tract (around 34 cm). It can be seen that transfer function for Model 2 is significantly impacted by gain adjustment in Itakura Distance Measure. As a result, as it can be seen in 3.32 in the legend and in Table 3.5 on the facing page, on average, second model performs better at modelling the vocal tract.

As tube model, we compared two models that are suggested in the section 2.3 on page 20. Similarly, in order to detect the vocal tract transfer function, we ignored the effect of the lip radiation and took reflection coefficient of glottis as 0.7. The additional parameters have been taken into account. We created spectra for four different excitation locations, mentioned in section 3.3.3 within the tube for both of the models. It is now possible to estimate an LP model for these eight spectra and use inverse filter on real snore events. However, since we do not know if these transfer functions do actually represent the transfer function for the airway during snoring sounds we compare the spectra with spectra obtained from IAIF model.

Snore event signals found in the overnight sleep recordings are collected and the middle frame from each snore signal has been extracted. The results from inverse filtering of each frame has been obtained and compared with eight transfer function spectra from the tube models using the distortion measures. As a result, distances to different models have been averaged, and the model with the minimum distance has been found. The results can be found in table 3.5 on the facing page

Patient	Average Itakura Distance and Deviation		Best Average Spectral Match			
	Model 1	Model 2	Model 1		Model 2	
			Type	Distance	Type	Distance
Patient 1	2.135 ± 0.564	0.714 ± 0.581	Epiglottis	1.396	Velum	0.376
Patient 2	1.989 ± 0.538	0.726 ± 0.487	Epiglottis	1.282	Velum	0.389
Patient 3	2.130 ± 0.558	0.659 ± 0.610	Epiglottis	1.581	Tongue	0.323
Patient 4	1.879 ± 0.549	0.761 ± 0.445	Epiglottis	1.139	Velum	0.374
Patient 5	2.099 ± 0.571	0.711 ± 0.591	Epiglottis	1.336	Velum	0.317
Patient 7	1.811 ± 0.541	0.713 ± 0.367	Epiglottis	1.129	Velum	0.410
Patient 8	2.104 ± 0.521	0.785 ± 0.554	Epiglottis	1.387	Velum	0.433
Patient 9	2.365 ± 0.595	0.666 ± 0.742	Epiglottis	1.399	Tongue	0.399
Patient 10	1.690 ± 0.514	0.844 ± 0.361	Epiglottis	1.017	Velum	0.436
Patient 11	2.082 ± 0.581	0.747 ± 0.560	Epiglottis	1.270	Velum	0.382
Patient 12	1.889 ± 0.594	0.691 ± 0.499	Epiglottis	1.063	Velum	0.265
Patient 13	2.028 ± 0.535	0.724 ± 0.472	Epiglottis	1.333	Velum	0.402
Patient 14	1.820 ± 0.522	0.728 ± 0.309	Epiglottis	1.246	Velum	0.469
Patient 15	2.294 ± 0.536	0.837 ± 0.577	Epiglottis	1.537	Tongue	0.374
Patient 16	2.189 ± 0.541	0.761 ± 0.549	Epiglottis	1.390	Tongue	0.409
Patient 17	2.195 ± 0.550	0.754 ± 0.601	Epiglottis	1.436	Tongue	0.423
Patient 18	2.210 ± 0.604	0.760 ± 0.684	Epiglottis	1.359	Velum	0.319
Patient 19	1.710 ± 0.459	0.759 ± 0.306	Epiglottis	1.105	Velum	0.456
Patient 22	1.960 ± 0.530	0.733 ± 0.484	Epiglottis	1.283	Velum	0.327
Patient 23	1.882 ± 0.572	0.796 ± 0.473	Epiglottis	1.087	Velum	0.356
Patient 24	2.010 ± 0.523	0.743 ± 0.508	Epiglottis	1.247	Velum	0.446
Patient 25	2.012 ± 0.538	0.702 ± 0.492	Epiglottis	1.320	Velum	0.374
Patient 26	1.914 ± 0.538	0.672 ± 0.468	Epiglottis	1.240	Velum	0.247
Patient 27	1.724 ± 0.438	0.804 ± 0.227	Epiglottis	1.332	Velum	0.430
Patient 28	1.820 ± 0.542	0.783 ± 0.417	Epiglottis	1.098	Velum	0.402
Patient 29	1.360 ± 0.360	0.997 ± 0.220	Epiglottis	0.876	Velum	0.699

Table 3.5: Average Spectral Distances and Standard Deviations for Model 1 and Model 2 to Vocal Tract Transfer Function Spectrum for each Patient

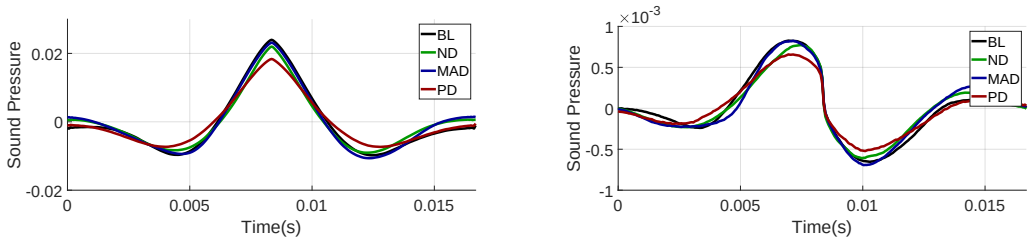
The averages are taken over all frames from various nights (including different device usage) and over 4 different hypothesized excitation locations for 2 models. It shall be noted that some devices may change the cross-sectional area of the airway during usage. We have also compared the distances for frequencies lower than 1500 Hz. As mentioned in the beginning of Chapter 3 the lower frequency region is dominates the energy spectrum and it is beneficial to compare these segments instead of the entire spectrum. In Table 3.5 it can be seen that average Itakura Distance to vocal tract transfer function spectrum for model two is significantly lower than Model 1. The standard deviations are taken over different excitation locations, and it can be seen that distances to Model 2 is highly sensitive to the assumed excitation location.

3.7. Analysis on Different Participants

Next, we have analyzed each patient and obtained excitation flow signal for each night and each snore epoch by averaging over all the snore events and interpolating and reshaping for median pitch period. The resulting signals can be seen below. This has been done for all the patients and it can be seen in Appendix B.

Patient	Model 1				Model 2				Sample Size
	Epiglottis	Tongue	Velum	O. Walls	Epiglottis	Tongue	Velum	O. Walls	
Patient 1	0.39	0	0.19	0.19	8.09	44.70	44.70	1.73	519
Patient 2	0.73	0	0	0.15	4.16	26.26	66.54	2.16	2041
Patient 3	0	0	0	0	11.76	23.53	64.71	0	17
Patient 4	4.81	0.69	0.69	0	1.10	22.01	68.09	2.61	727
Patient 5	0.82	0	0	0	5.56	32.10	59.67	1.85	486
Patient 7	3.23	0.92	0	0.46	4.15	15.21	71.89	4.15	217
Patient 8	0.67	0.22	0	0.11	3.14	40.31	49.27	6.27	893
Patient 9	0	0	0	0	60.38	18.87	20.75	0	53
Patient 10	7.90	2.83	1.30	0.24	0.94	6.84	71.58	8.37	848
Patient 11	6.43	0	0	0.23	20.42	17.93	52.87	2.12	1322
Patient 12	0.93	0	0	0	0.31	7.41	91.05	0.31	324
Patient 13	0	0	0	0	0	46.15	53.85	0	26
Patient 14	0	0	0	0	0	21.88	75	3.13	32
Patient 15	0	0	0	0	28.95	28.07	41.23	1.75	114
Patient 16	0	0	0	0	16.25	40	41.25	2.50	80
Patient 17	2.42	0.40	0	0	32.66	26.21	33.87	4.44	248
Patient 18	0.41	0	0	0	1.51	32.74	65.21	0.14	730
Patient 19	0.76	1.19	0.11	0.32	0.43	15.43	69.69	12.08	927
Patient 22	0.68	0.68	0	0	0	30.82	65.75	2.05	146
Patient 23	5.90	0.88	0.14	0.47	1.15	10.52	77.41	3.53	1474
Patient 24	0	0.56	0.56	0	0.56	31.07	63.84	3.39	177
Patient 25	0	0	0	0.64	0.64	38.46	59.62	0.64	156
Patient 26	0	0	0	0	0	9.09	90.91	0	11
Patient 27	0	0	0	0	0	16.22	59.46	24.32	37
Patient 28	6.70	0.90	0.05	0.42	0.42	15.41	69.55	6.54	1895
Patient 29	17.94	9.16	4.58	0	0	17.18	29.77	21.37	262

Table 3.6: Percentage for detected snoring instances per obstruction type for both models



(a) Excitation Waveform signal for patient 1

(b) Excitation Derivative Waveform for patient 1

Figure 3.33: Waveforms for patients. Each line indicates a snore epoch. Bold lines indicate averaged flow signals for each device

We select specific patients that have been identified as whose snoring has been substantially alleviated by a device, as mentioned in subsection 3.5.1 on page 43. Here, we will demonstrate each flow signal obtained from snore epochs of Baseline duration and try to identify if there is a significant difference in the excitation flow between the set of patients impacted by Nasal Dilator, Mandibular Advancement Device, and Positional Device. Our expectation is to find different flow signals and different pitch for group of patients, as the patients reported that this specific device remarkably changes their snore pattern. As a first investigation, we look at the number of snore events detected using our YIN estimator.

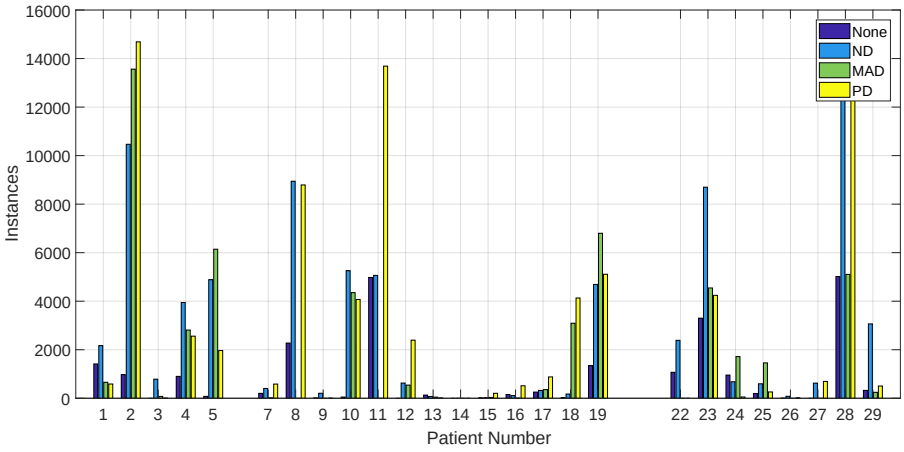


Figure 3.34: Total Events per Device

We also calculate average number of events.

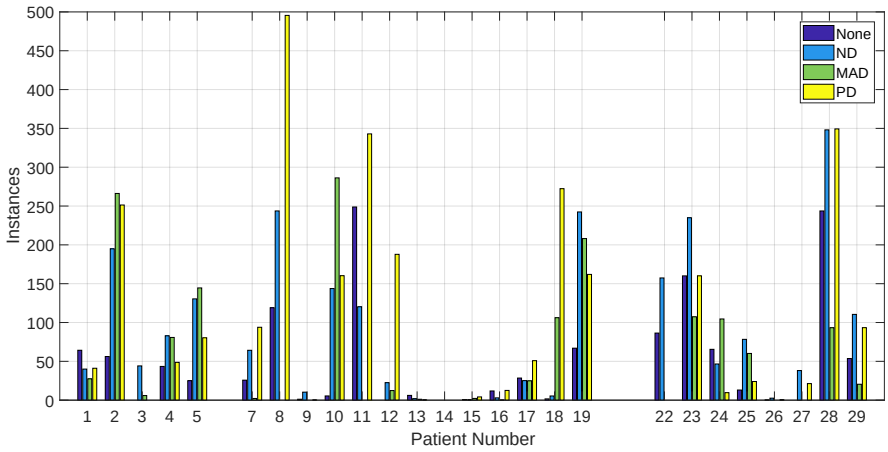


Figure 3.35: Average snore events per device per hour

Next, we show the results for each excitation flow analyzed from the snoring signals. We analyzed the snore epochs extracted from the sleep periods during the baseline period of first three days. The participants for whom specific devices worked particularly well as identified in Brinker et al. [19] can be seen in Table 3.7 on the next page.

Effective Device	Participant	# of Nights	# of snore epochs
Nasal Dilator	4	3	8
	5	3	1
	6	3	1
Mandibular Advancement Device	3	2	0
	15	3	12
	18	3	6
	22	3	4
	24	3	5
	26	2	1
	27	3	0
	28	3	22
Positional Device	7	3	1
	9	3	1
	14	3	0
	19	3	6
	25	3	3

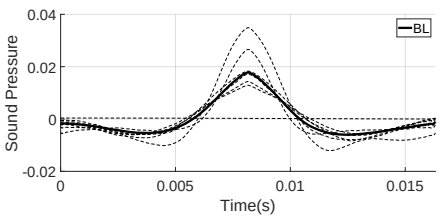
Table 3.7: Effective Device Data From the trial

As it can be seen in table 3.7 above several participants lack sufficient knowledge for baseline period. Participant 3 and Participant 14 are examples of this case. Additionally, Participant 9 has one baseline period recording which features a high background noise.

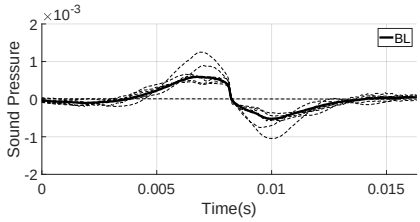
Moreover if we compare the results in Figure 3.35 on the preceding page and 3.34 with Table 3.7, we see that reported effective devices do not reflect a decrease in number of snore events per hour or number of snore events. This result might point to different conclusions. Moreover, the reporting of the partners may have different biases and different priorities for characterizing sleep quality. These will be elaborated deeply in the discussion chapter.

Below, we introduced excitation flow waveforms extracted from sound files during the baseline duration of the trial, which refers to the duration that no device has been used. We aim to find a distinct pattern for patients which a different device worked well. We average the excitation flows of each snore event, and then average over the entire snore epoch. In figures below, dotted lines show each night, and solid line shows the average.

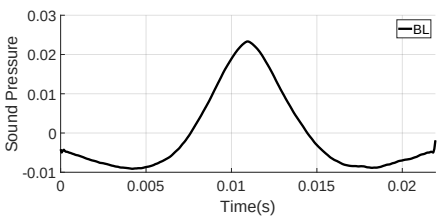
Participants Alleviated by Nasal Dilator (ND)



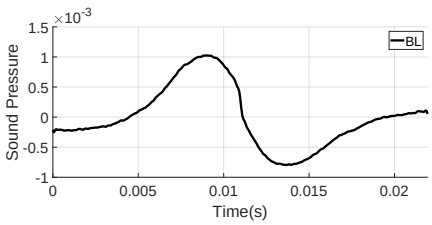
(a) Patient 4 Excitation Flow



(b) Patient 4 Excitation Flow Derivative

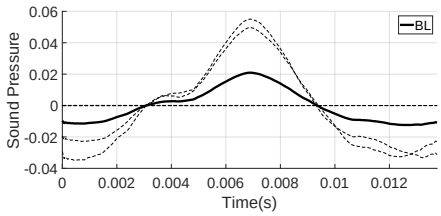


(a) Patient 5 Excitation Flow

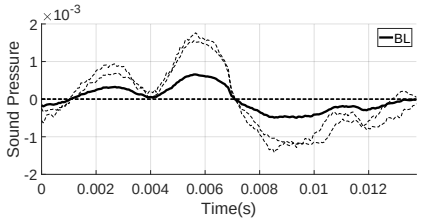


(b) Patient 5 Excitation Flow Derivative

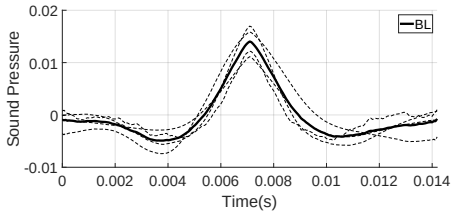
Participants Alleviated by Mandibular Advancement Device (MAD)



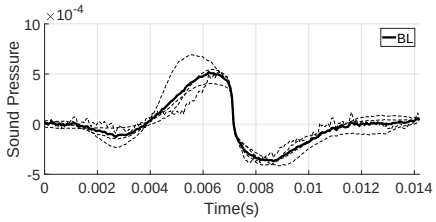
(a) Patient 18 Excitation Flow



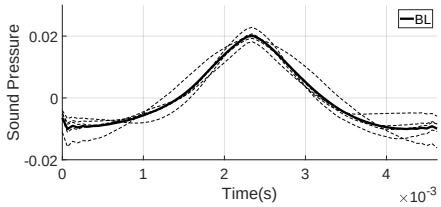
(b) Patient 18 Excitation Flow Derivative



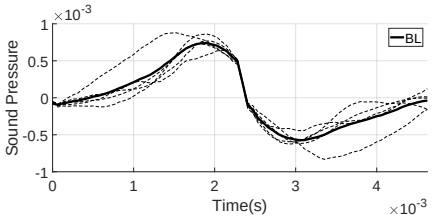
(a) Patient 22 Excitation Flow



(b) Patient 22 Excitation Flow Derivative

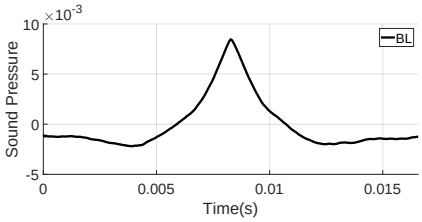


(a) Patient 24 Excitation Flow

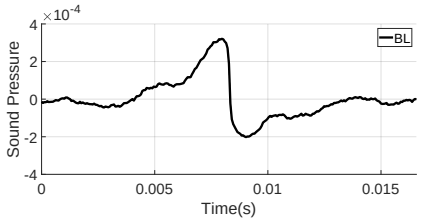


(b) Patient 24 Excitation Flow Derivative

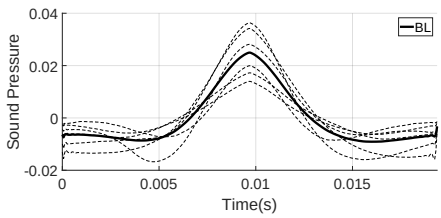
Participants Alleviated by Positional Device (PD)



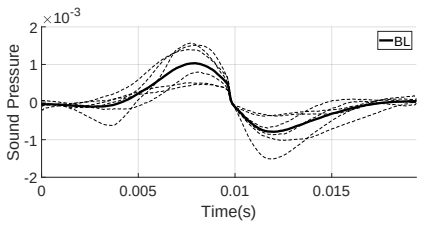
(a) Patient 7 Excitation Flow



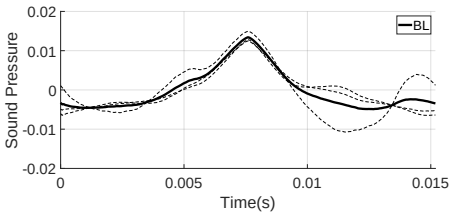
(b) Patient 7 Excitation Flow Derivative



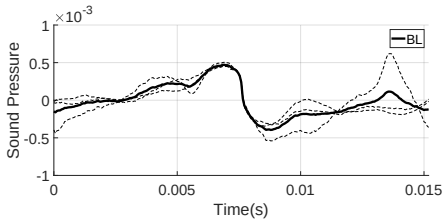
(a) Patient 19 Excitation Flow



(b) Patient 19 Excitation Flow Derivative



(a) Patient 25 Excitation Flow



(b) Patient 25 Excitation Flow Derivative

Resulting waveforms are discussed in Chapter 4 on the facing page.

4

Discussion

The experiments section showed us several aspects of tube modelling and whether these can be applied to the snoring problem. In this section we will elaborate the results of these experiments.

Unlike Ng et al.[64] and Saha et al.[73], we have followed different methods and procedures to work on the upper airway. Ng, excited the tube from glottis, and used a transfer function created for the entire upper airway tube. As an hypotheses we suggested alternative models, which have different excitation locations and different circuit characteristics in section 2.3. Additionally, in section 3.3 we have shown that obstruction location plays a critical role in creating the upper airway transfer function for the models that have been suggested. However, similar to Ng, we have found that narrowing of the pharyngeal airway increase the formant frequencies for Model 1 and peak frequency, as expected from the tube modelling theory.

On the other hand, Saha et al. used tracheal recordings, which differ from ambient recordings and claimed to be higher in SNR and useful in classification accuracy [84]. Additionally, they have only experimented on non-varying cross sectional area, i.e. single tube. In their studies they have found that increases in the upper airway length reduced the resonant frequency of snoring sounds and validated against recordings of actual snoring events. This also matches with our findings for model one and model 2 in Figure 3.13 on page 37 and 3.14 on page 37.

Apneic(OSA) and benign(non-OSA) snoring sounds have been compared in Fiz et al. [36], Perez-Padilla et al. [67]. While the results do not reach a consensus on difference between OSA and non-OSA patients, it has been found that OSA snorers have power spread over higher frequency range compared to non-OSA snorers. Additionally it has been found that maximum frequency i.e upper limit of the spectrum defined as the highest frequency peak with a power greater than 3% of the peak power has been significantly larger(500Hz non-OSA, 1400Hz OSA) in OSA patients with a larger spread. This can be linked to our study in the sense that if the constriction is in multiple regions and in the upper part of the airway, Oropharyngeal walls, is the area that is one of these areas for both of the models, the first formant frequency increased. Especially, as seen in 3.14 on page 37 first formant frequency is around 1200 Hz. Hence, it may be claimed that this increase in the first formant frequency may reflect characteristics of multiple constrictions as in OSA. However, since OSA patients have been excluded from the trial, it would be overreaching to conclude on characteristics of OSA snore events.

Additionally, previous studies have found that main frequencies of snoring due to the narrowing at the base of tongue were above 650 Hz, frequencies of palatal snoring were below 450 Hz [83][70]. This has also been observed in our models, specifically in Model 2 3.14 on page 37.

As the next step, we discuss the characteristics of the excitation flow. Firstly, we observed

that speech signals and snore signals have fundamentally different flow characteristics as it can be seen in Figures 3.16 on page 38 and 3.21 on page 41. This can be observed using the inverse filtering technique that has been used for speech processing for snoring signals and speech signals. This difference is expected as speech is created from forced vibration of the vocal folds, while snoring is caused by vibration of the relaxed tissue that is obstructing the airway. We also found that pitch of the snoring signals are lower compared to speech signals for most of the participants to the trial.

Snoring excitation waveforms may differ for each person and each device used during the sleep as seen in Section 3.7 on page 51. It also varies for each snore event and each night. However, they do follow a similar pattern and this enables the parametrization of excitation flow [63]. The waveform follows an exponential rise for the first half of the flow pattern, and a fall, again exponentially. While there exists difference between each device period waveforms, it is difficult to achieve a consensus on how each device effects the excitation waveform. The difficulty comes from varying pitch during snoring, lack of audio files for many of the patients for many of the nights, and noisy sound files. Additionally it is hard to achieve consensus on what creates a difference between participants during the baseline period. Parametrization of these snore flows is a plausible way to find the differences in these flow signals. Airas et al.[3], and Ng[63] defined multiple parameters for glottal waveform, namely, Open Quotient, Negative Quotient, Positive Quotient etc. Ng et al. [63] has claimed that these features are useful for differentiating snore signals.

4.1. Limitations

In this section we will mention to limitations of the experiments and mention different ways to surpass them.

As reported in [69], [53], [54] the obstruction sites might be plural. While our model is capable of simulating such a case, inverse filtering methods would not be able to find multiple snore flows for different locations. The models can be adjusted to excite the tube from multiple locations with different excitation flows. However, in such a case, the transfer function calculation should be adapted for multiple sources.

While nasal patency has been in the scope of this problem, it has not been investigated for this study. Adding the impact of a closed or open nasal opening is trivial for both of the models using a parallel input parallel output branching near the volume area as mentioned in Maeda [56]. However, due to time schedule this is left for future work.

Effect of turbulence has not been investigated for this study. Reynolds number has to be calculated, to determine if the flow characteristic is laminar or turbulent. If we calculate the Reynolds number using mean inhalation velocity associated with sinusoidal at-rest [7], $1.81\,\mathrm{m\,s^{-1}}$, air density, ρ as $1.14\,\mathrm{kg\,m^{-3}}$ and diameter as $5\times10^{-2}\,\mathrm{m}$ and dynamic viscosity as $1.85\times10^{-5}\,\mathrm{Pa\,s}$ we find the Reynolds number between 450 for 4cm and 670 for 6 cm. For flow in a pipe of diameter D, Reynolds number below 2300 points to laminar flow [74]. However, changes in cross-sectional area, and in turn, changes of the volume velocity within the tube with respect to depth of the airway, might cause turbulence during inhalation. This has also been pointed out in Quinn et al. [70] as tongue base snoring appears to be associated with airflow turbulence in a rapid, continuous airflow through a severely narrowed airway. These turbulence-induced sounds have a higher frequency and broader spectral range than that produced by the floppy soft palate intermittently obstructing the nasopharyngeal airway. Detailed research on fluid dynamics using Finite Element Methods, and Computation Fluid Dynamics have been conducted by Balint et al. [11] Chouly et al. [26], Bertram [14], Zhao et al. [85] and Lucey et al. [55] and is an active field in fluid dynamics.

Total duration measured for each night is not precise. Audiovisual inspection showed that many participants do not immediately fall asleep. However, as it is hard to point to the time instance that patient had begun sleeping, it is incrementally hard to find the exact number of snore events per hour. This, effects the results in 3.35 on page 53 and 3.34 on page 53.

Additionally, as we use a Voice Activity Detector, we also detect surrounding sounds caused by TV, crying baby, door opening, etc. Even though we have used a temporal limit of 2 seconds for detection of inhalation events to eliminate faulty detections, this does not overcome the interference problem for accurately detected segments. This overlap with snoring sounds, causes the spectrum to be polluted by environmental sounds.

Different devices do change the waveform of the excitation flow. However, it is hard to deduce results from this trial. Number of patients and the non-clinical setup of the trial makes the assessment of the flow waveforms intractable. As mentioned in the previous sections, creation of the ground truth data is mostly derived from assessment of the participants', and their partners' reports in the morning after and reporting at the end of device trial period. As these reports can be susceptible to many bias problems such as Consistency bias, Selective Memory, Self-Serving Bias, Rosy retrospection, Anchoring , Attentional bias and Focusing effect and Peak-end rule.

The assumption that flow and the airway tract are linearly separable in section 2.5 on page 24 may not be true for snoring problem. As mentioned in Babbs [10], Venturi effect plays an important role in snoring. As Venturi effect is defined as reduced pressure that occurs when a fluid flows through a constricted section of a tube or pipe, it changes to cross sectional area. This means that tube and the flow are not linearly separable.

5

Conclusion and Future Work

The thesis project is aimed at designing an unobtrusive method to find the obstruction location for benign(non-OSA) snoring patients, during their non-sedated sleep using simple recording devices within uncontrolled environment. The method that has been suggested has been created with an aim to reflect the physical realities of sound production.

We considered tube modelling in order to model snoring sound production. In the first part of the thesis we have explained how our tube model works and created two different models in concordance with upper airway anatomy. Additionally, we have introduced additional components that reflect the physical phenomena, such as walls, heat conductance, and resistance and compared its impact on tube modelling in experiments chapter. We also investigated if linear prediction can be used to determine the shape of the upper airway tract, and in concordance with the tube model. Later we tried to determine if these two methods, namely tube model, and linear prediction can help to determine the excitation flow during snoring. This has been later tested on snoring sounds recorded in natural conditions. While the method has not been fully assessed and proven, due to numerous reasons (presented in Discussion), it has been shown that it is in concordance with previous clinical trials and research.

The concluding remarks on this study can be found below

1. Remarks on Tube Modelling

   (a) We checked transfer function spectra for different models and different excitation locations and obstruction severities. We have observed that for Model 1 the first formant of the transfer function has increased as we moved the excitation location to upper regions of the airway, equivalently, creating a shorter tube. For model 2, the results show that it is harder to derive direct conclusions as model 1.

   (b) Additional components do not have a significant impact on reconstruction using linear prediction coefficients, as these components do not alter the transfer function spectrum of the vocal tract.

   (c) Tube modelling using transfer function for ducts can model the upper airway section while vowels are being pronounced. Tube model do find the same peaks and bandwidths compared to commonly used inverse filter models. However, it is important to point the determination of the order of the linear prediction suffers from the problem of causality dilemma, as the obstruction location is not known. This especially has a higher impact on model two, as the lower part of the respiratory tract alters the tube model and the physical rules for determining the model order can not be applied for this case.

   (d) Model 2, which incorporates the lower part of the airway tract finds lower formant frequencies around 300-500 Hz, which is parallel to previous literature.

2. Remarks on Linear Prediction and Inverse Filtering

   (a) Investigating viability of linear prediction for estimating the upper airway has been done using different experiments. Based on these test, you only get a correspondence on certain settings on a specific model and certain conditions, with specific wall parameters, specific termination. Therefore, it is necessary to point that it is questionable to use linear prediction, since under natural conditions, not all the requirements are met.

   (b) We have shown that using linear prediction coefficients to reconstruct the upper airway is not a plausible method as the spectral distances between vocal tract transfer function obtained from the inverse filter and transfer function obtained from the tube model are large. As we have confirmed that LPC can only represent a model with an anechoic ending at excitation location in Model 1, using the coefficients to reconstruct the cross-sectional area is not a plausible method. Consequentially, the estimated transfer function from the inverse filter do not represent the cross-sectional area.

   (c) Even though Linear Prediction is not a good method to get the cross-sectional areas, it is useful to find the temporal characteristics of the snore signal i.e excitation flow, and spectral characteristics, i.e. upper airway transfer function.

   (d) An excitation flow pattern for snoring has been found. This is similar to previous literature [47], but it is yet to be confirmed with better imaging techniques.

3. Remarks on Trial Data and Phenotyping

   (a) Vocal tract transfer functions from inverse filtering methods have been compared with the transfer functions of models excited from different locations. Using the spectral distance measures, it has been found that, Model 2 is more successful in modelling the airway tract for snoring problem.

   (b) For Model 1, the majority of the airway tract transfer functions has been identified as Epiglottic Snores, while for Model 2 the identified snores have been diverse.

   (c) The effectiveness of the devices can not be identified by simple investigation of average excitation flow, and we have not found any convincing leads that there is a difference between flow models regarding the shape. Further work needs to be conducted on parametrization to discover a distinguishing factor, given that subjective evaluations of the participants reflect the essence of the problem accurately. Additionally, it is important to detect the pitch period accurately, in order to envelop the excitation flow signals correctly.

   (d) Average number of snores per hour do not reflect the reporting of the participants regarding the efficient device. While loudness can be an important aspect, it can be said that annoyance of the partners are very subjective, and hard to estimate using scientific methods.

5.1. Future Work

In order to achieve the main goal, presenting a product that correctly finds the obstruction location, and suggests a device for the consumer, several steps still need to be taken. These steps are as follows

1. Annotation of the obstruction and excitation location has to be made by Ear Nose Throat (ENT) experts similar to the previous studies. As we do not know the source of the location that we have used in our analysis, this still need further validation by medical personnel.

2. A much better validation could be achieved by using a discriminative measure suggested by Robb et al. [72]. This would require the patients withing this study to pronounce multiple vowel sounds to predict snoring propensity. This would enable the method to estimate the cross-sectional area and the constrictions within the airway tube before recording the snoring sounds.

3. To the best of our knowledge, excitation flow pattern for snoring is yet to be confirmed by advanced imaging techniques. Further, a controlled trial, using better audio equipment can be used to measure the loss of quality and accuracy using mobile phones.

4. A mobile application, that can easily be used by consumers and give sufficient report on how they sleep throughout the night has to be developed.

A

IAIF vs DIF Comparison

Comparison of flow models from TIMIT database for specified vowels.

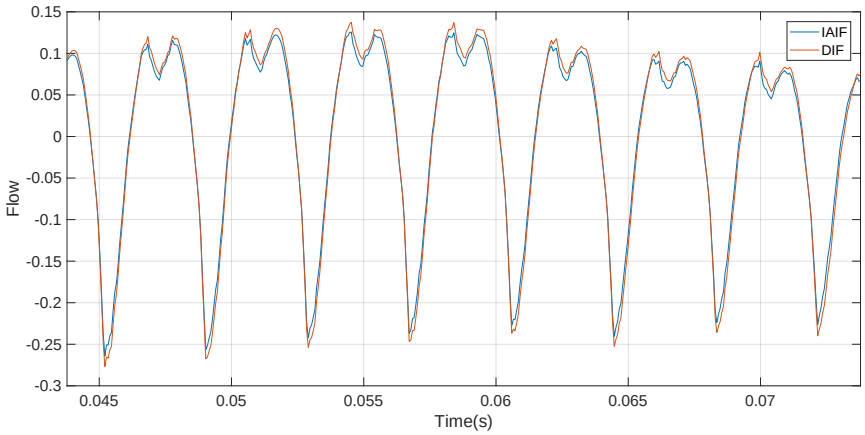


Figure A.1: Inverse Filtering Model Comparison for ɛ

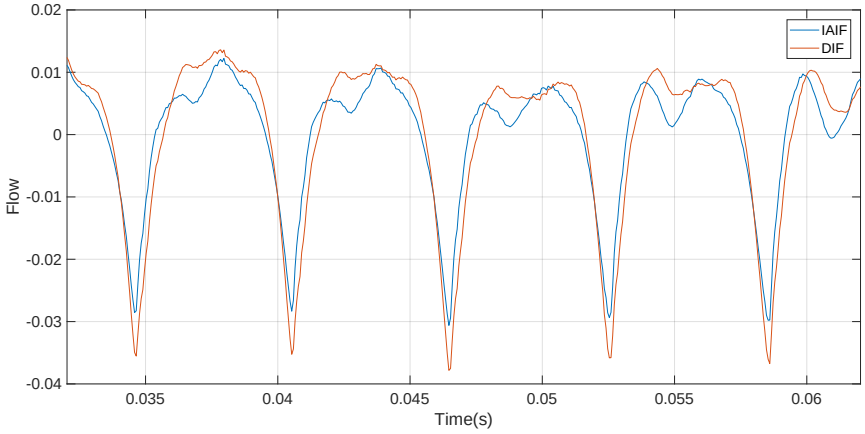


Figure A.2: Inverse Filtering Model Comparison for ɪ

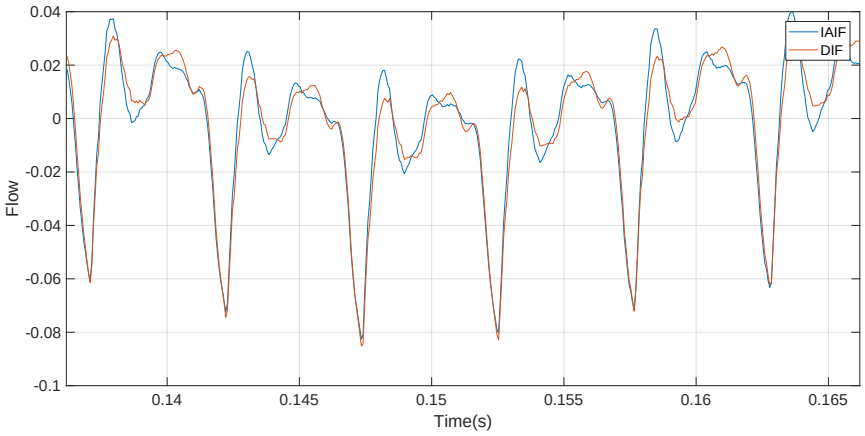


Figure A.3: Inverse Filtering Model Comparison for ɔ

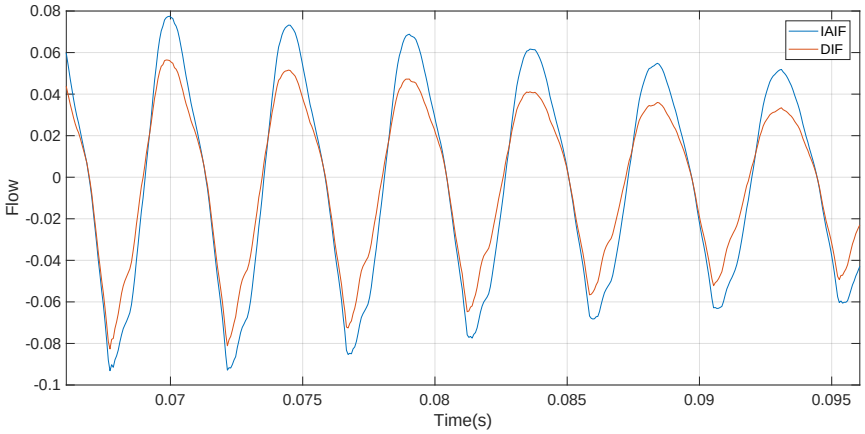


Figure A.4: Inverse Filtering Model Comparison for u

B

Comparison of Excitation Flow for Different Device Periods

The excitation flow for different device periods for each patient is shown in this section. If a flow does not exist it might have different meanings:"

1. The patient has not made successful recordings for the device period
2. The patient has not filled in the survey regarding which device he/she used.
3. The patient did not snore on that night.
4. The patient did snore, but the voice activity detector has not captured it.

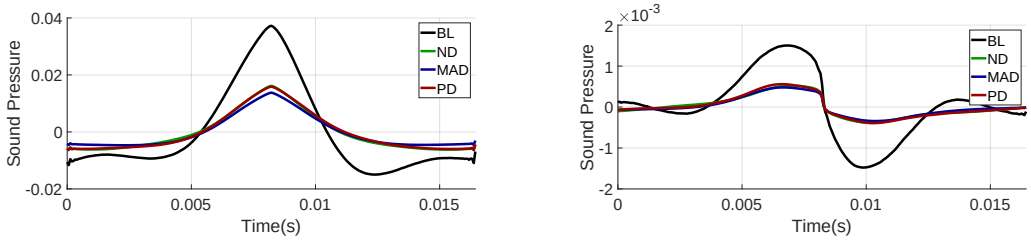


Figure B.1: Excitation Waveform(left) and Excitation Derivative Waveform(right) for patient 2

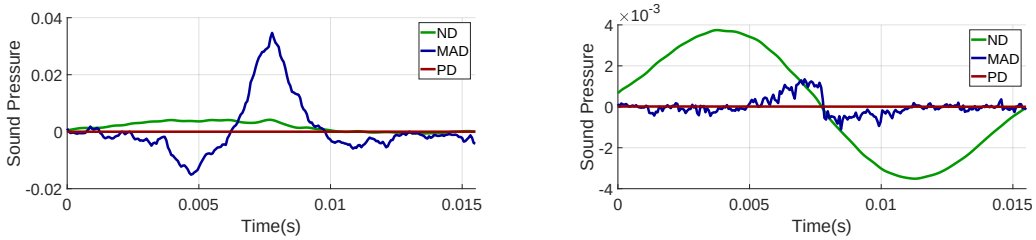


Figure B.2: Excitation Waveform(left) and Excitation Derivative Waveform(right) for patient 3

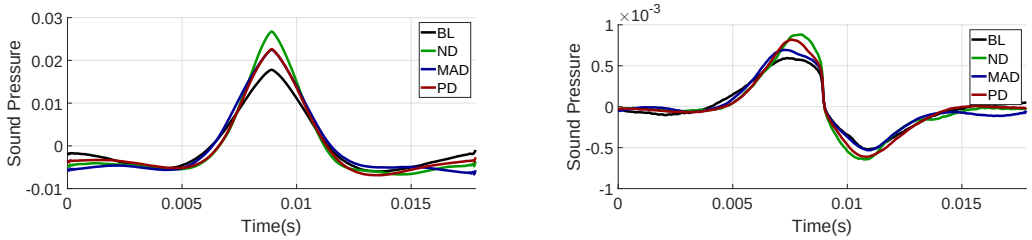


Figure B.3: Excitation Waveform(left) and Excitation Derivative Waveform(right) for patient 4

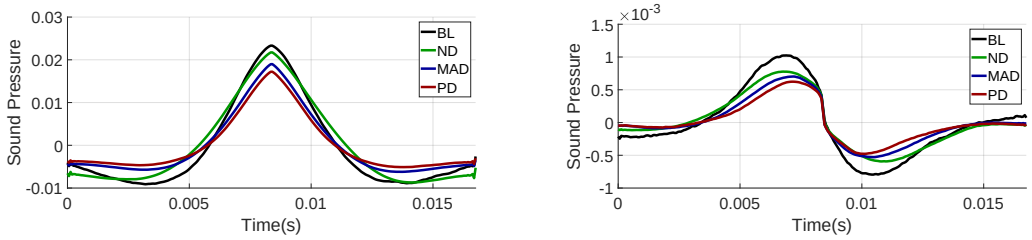


Figure B.4: Excitation Waveform(left) and Excitation Derivative Waveform(right) for patient 5

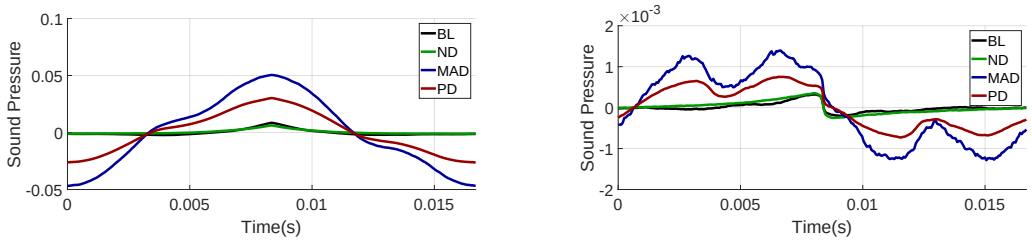


Figure B.5: Excitation Waveform(left) and Excitation Derivative Waveform(right) for patient 7

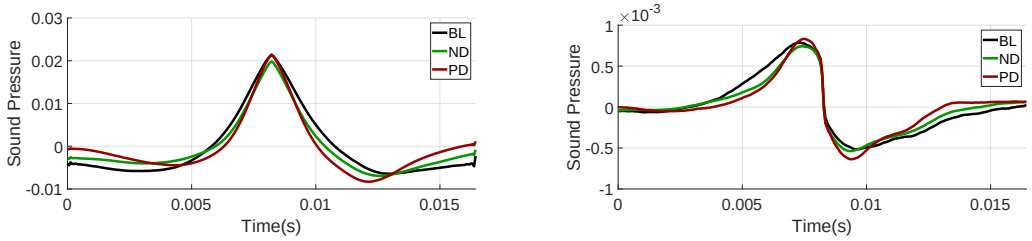


Figure B.6: Excitation Waveform(left) and Excitation Derivative Waveform(right) for patient 8

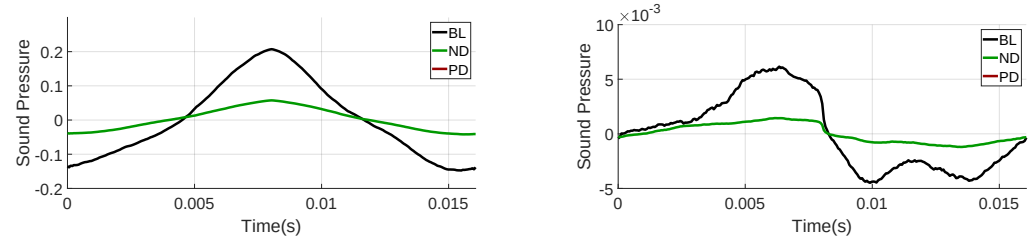


Figure B.7: Excitation Waveform(left) and Excitation Derivative Waveform(right) for patient 9

C

Vowel Transfer Function Comparison

This appendix section shows the comparison for transfer functions obtained from sound files using inverse filtering, and obtained from cross-sectional area tube model transfer functions.

E

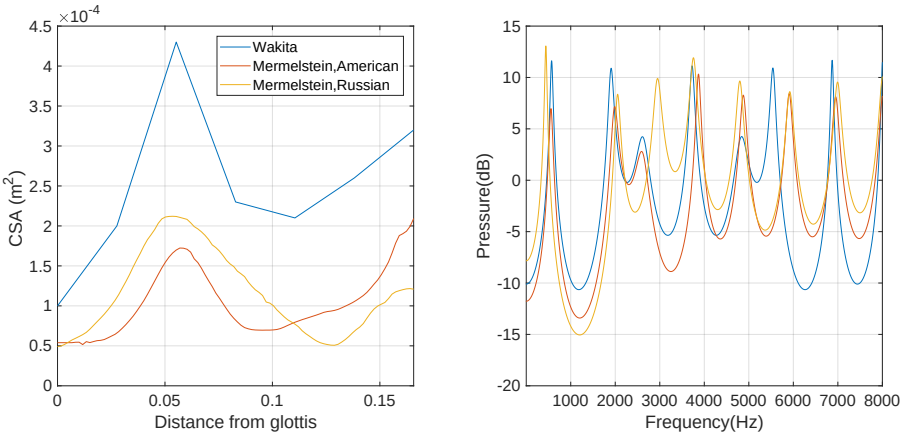


Figure C.1: (left) Cross-sectional Areas; (right) Transfer Function for vowel ε

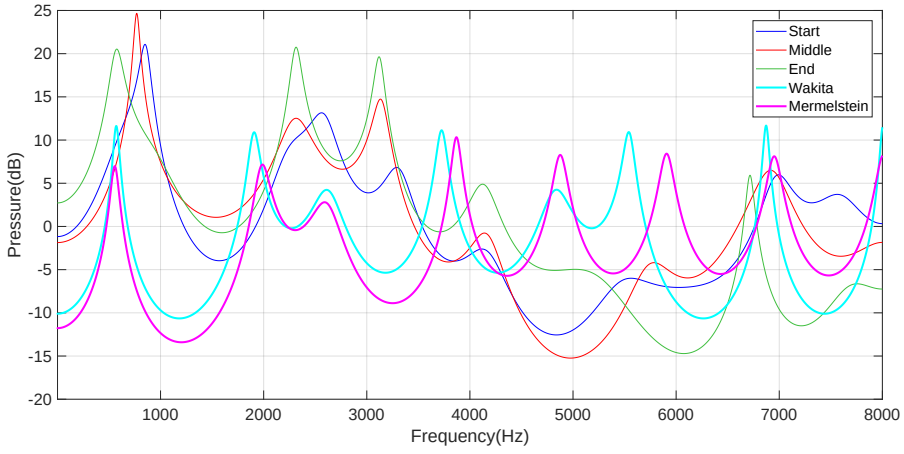


Figure C.2: Comparison of Vocal Tract Transfer functions from Inverse Filter vs Tube Model for vowel ε

ɪ

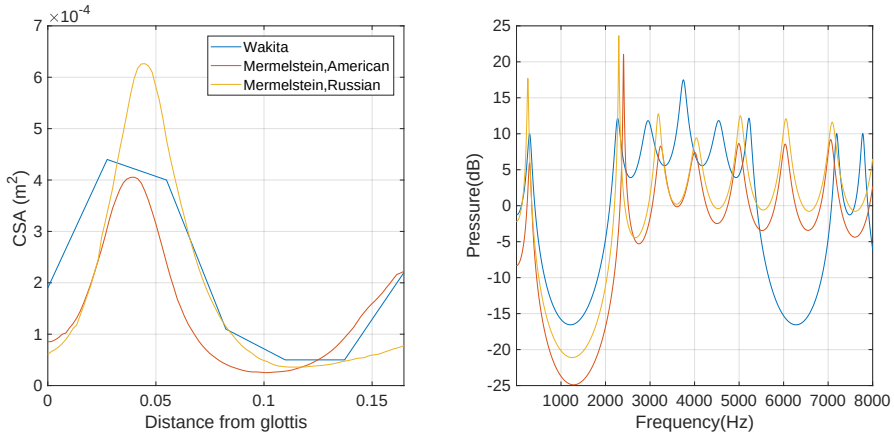


Figure C.3: (left) Cross-sectional Areas; (right) Transfer Function for vowel ɪ

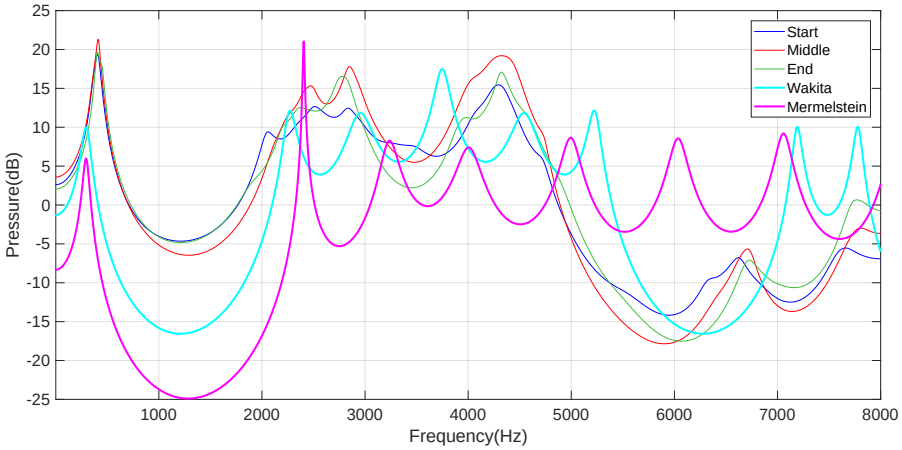


Figure C.4: Comparison of Vocal Tract Transfer functions from Inverse Filter vs Tube Model for vowel ɪ

ɔ

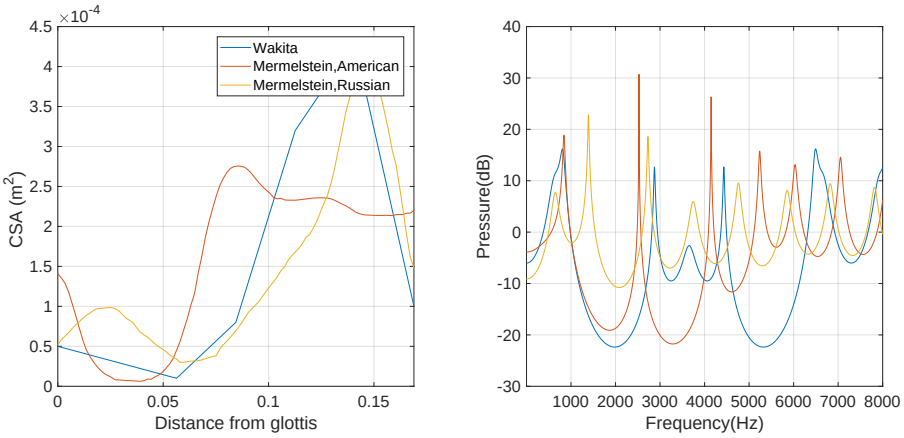


Figure C.5: (left) Cross-sectional Areas; (right) Transfer Function for vowel ɔ

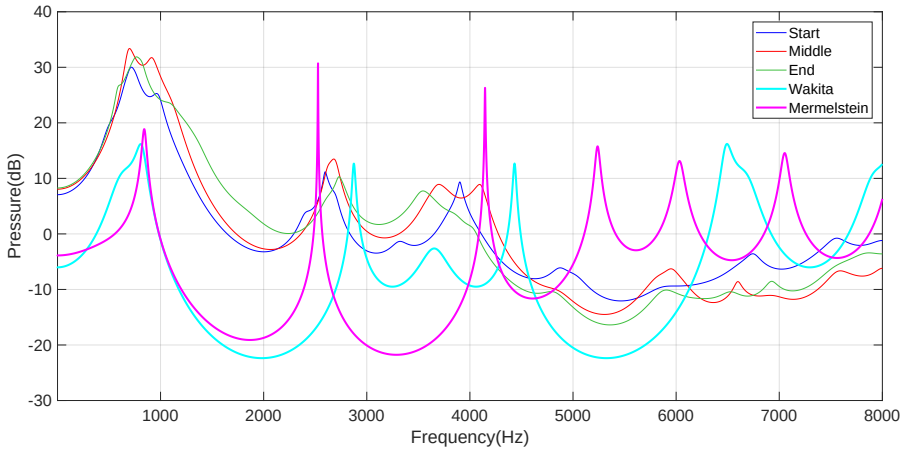


Figure C.6: Comparison of Vocal Tract Transfer functions from Inverse Filter vs Tube Model for vowel ɔ

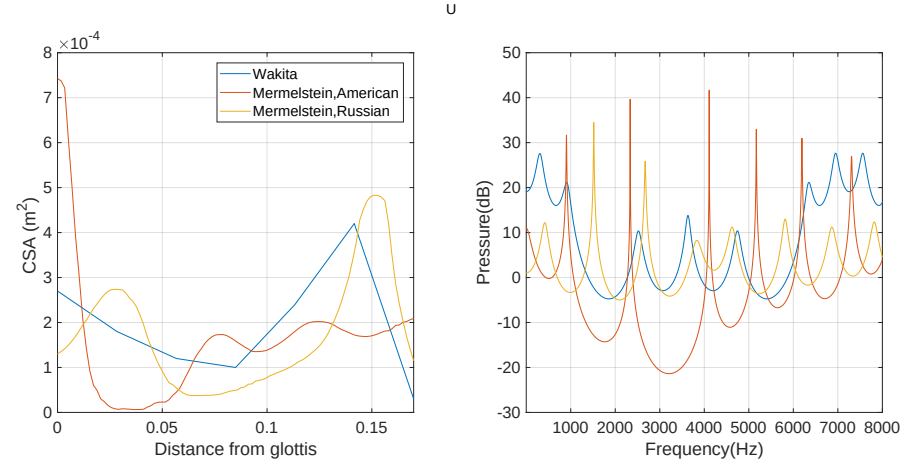


Figure C.7: (left) Cross-sectional Areas; (right) Transfer Function for vowel ʊ

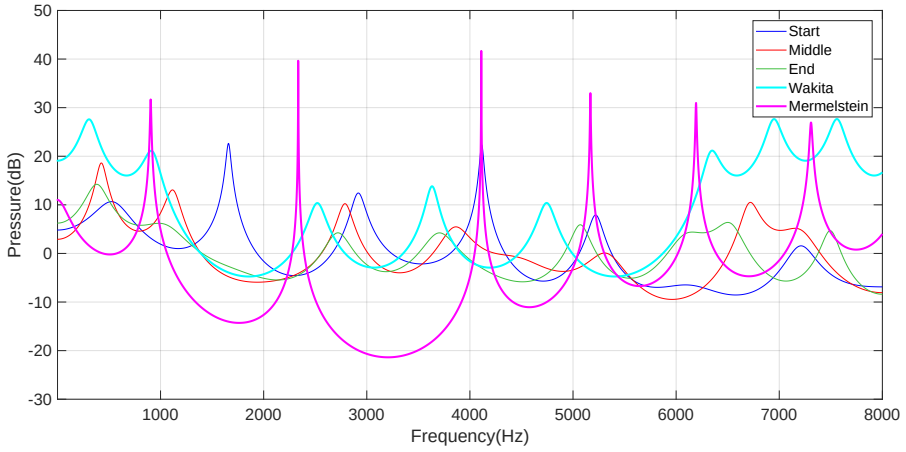


Figure C.8: Comparison of Vocal Tract Transfer functions from Inverse Filter vs Tube Model for vowel ʊ

D

Software Framework

This chapter will describe the software that has been written to enable the research. All of the files are included in subdirectory *proteus/snoringmodel*

Directory / Name	Description
areameasurements / measurements	Cross-Sectional Area Measurements digitized from [60] and [28]
areameasurements / measurementsletter-American	Cross-Sectional Area Measurements for American English vowels digitized from Mermelstein et al. [58]
areameasurements / measurementsletter-Russian	Cross-Sectional Area Measurements for Russian vowels digitized from Mermelstein et al. [58]
areameasurements / CSANarrow	Narrowing cross sectional area calculation The method does interpolation for equally distributed pivot points, that has been narrowed proportionally.Function that finds a narrower cross sectional area calculation given the parameters obstruction index,obstruction severity ,obstruction width
test / testWallParams	Goes over each wall parameters in Ishizaka et al. and Harper et al. Plots the output transfer function and reconstruction
test/ testlowerrespiratoy	function that gives impedance for lower respiratory system
test / spectral_inspection	Function that creates different cross-sectional areas and excitation locations, and plot transfer function and Cross-Sectional area reconstructions based on Model 1
test / spectral_inspection_end	Function that creates different cross-sectional areas and excitation locations, and plot transfer function and Cross-Sectional area reconstructions based on Model 2
multitube_source_location	Function that builds the tube model
multitube_source_location_end	same script with multitube_source_location with the exception that first half of the tube is taken into account and an anechoic termination is added - Model2
average_excitation_flow	Finds average excitation flow for voiced segments using inverse filtering methods.

find_wform	Function that finds the maximum or minimum and encapsulates the signal with pitch period.
test / testR	Function that test impact of Resistance on the transfer function and CSA reconstruction
test / testG	Function that test impact of Conductance on the transfer function and CSA reconstruction
distance_spectrum_table	Function that outputs tables for distance of transfer functions of inverse filter and tube model.
test / testYin	Test function on snoring signals for Yin Pitch Estimation
test / plotSnoreCount	Function that takes patient information and outputs snore count in a histogram plot
test / plotPatientWaveform	Function that takes patient information and outputs snore count in a histogram plot
test / plotEffectiveWform	Function that reads patient data and finds the average flow for each patients baseline periods
test / spectral_inspection_VOTE	Function that creates different cross-sectional areas and excitation locations, similar to the profiles described in Kezirian et al.(2011) and plot transfer function and Cross-Sectional area reconstructions.
test / spectral_inspection_RG	Function that compares outputs for varying glottis reflection coefficient
AnimatedAcousticTube /atvframe	Function that calculates the Areas belonging to the acoustic tube needed to generate the sound defined by the variable input.
airwaywall	struct that hold the parameters for airway wall.

Glossary

Anchoring he tendency to rely too heavily, or "anchor," on a past reference or on one trait or piece of information when making decisions.. 59

Attentional bias neglect of relevant data when making judgments of a correlation or association.. 59

Consistency bias incorrectly remembering one's past attitudes and behaviour as resembling present attitudes and behaviour. . 59

Focusing effect prediction bias occurring when people place too much importance on one aspect of an event; causes error in accurately predicting the utility of a future outcome.. 59

Peak-end rule that people seem to perceive not the sum of an experience but the average of how it was at its peak (e.g. pleasant or unpleasant) and how it ended.. 59

Rosy retrospection the tendency to rate past events more positively than they had actually rated them when the event occurred.. 59

Selective Memory refers to rare events which are widely reported, thereby altering the perception of how common they actually are. . 59

Self-Serving Bias the tendency to attribute successes to internal characteristics while blaming failures on outside forces. It may also manifest itself as a tendency for people to evaluate ambiguous information in a way beneficial to their interests.. 59

Acronyms

AHI Apnea-Hypoapnea Index. 4

AR Autoregressive. 22

BL Baseline Period. x, 44

BPM Beats Per Minute. 44

DIF Direct Inverse Filtering. 24, 38

DISE Drug Induced Sleep Endoscopy. 4

ENT Ear Nose Throat. 3, 4, 62

FIR Finite Impulse Response. 24

IAIF Iterative Adaptive Inverse Filtering. 24, 38, 50

IIR Infinite Impulse Response. 22

LP Linear Prediction. 35, 36, 39

LPC Linear Prediction Coefficients. 32, 62

LRT Likelihood Ratio Testing. 4

MAD Mandibular Advancement Device. x, 44, 52, 55

MPSSC Munich-Passau Snore Sound Corpus. 4

ND Nasal Dilator. x, 44, 52, 54

NREM Non-Rapid Eye Movement. 2

OSA Obstructive Sleep Apnea. 1, 3, 43, 57, 61

PCA Principal Component Analysis. 3

PD Positional Device. x, 44, 52, 55

RDI Respiratory Disturbance Index, sum of apneas and hypopneas. 27

REM Rapid Eye Movement. 2

ROC Receiver operating characteristic. 4

SFM Source Filter Model. 5, 42

SNR Sound to Noise Ratio. 4, 40, 47, 57

Bibliography

[1] R. M. Aarts and A. J. E. M. Janssen. Approximation of the struve function h1 occurring in impedance calculations. *The Journal of the Acoustical Society of America*, 113:2635–2637, 2003. ISSN 0001-4966. doi: 10.1121/1.1564019.

[2] S. Agrawal, P. Stone, K. McGuinness, J. Morris, and A. E. Camilleri. Sound frequency analysis and the site of snoring in natural and induced sleep. *Clinical Otolaryngology*, 27(3):162–166, 2002.

[3] M. Airas. TKK Aparat: An environment for voice inverse filtering and parameterization. *Logopedics Phoniatrics Vocology*, 33(1):49–64, 2008. ISSN 14015439. doi: 10.1080/14015430701855333.

[4] T. Aittokallio, M. Gyllenberg, and O. Polo. A model of a snorer's upper airway. *Mathematical Biosciences*, 170(1):79–90, 2001. ISSN 00255564. doi: 10.1016/S0025-5564(00)00062-6.

[5] H. Alshaer, M. Garcia, M. H. Radfar, G. R. Fernie, and T. D. Bradley. Detection of upper airway narrowing via classification of LPC coefficients: Implications for obstructive sleep apnea diagnosis. In *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, pages 681–684. IEEE, 2011.

[6] H. Alshaer, A. Pandya, T. D. Bradley, and F. Rudzicz. Subject independent identification of breath sounds components using multiple classifiers. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, pages 3577–3581, 2014. ISSN 15206149. doi: 10.1109/ICASSP.2014.6854267.

[7] T. R. Anthony and K. R. Anderson. Computational fluid dynamics investigation of human aspiration in low-velocity air: orientation effects on mouth-breathing simulations. *Annals of occupational hygiene*, 57(6):740–757, 2013.

[8] A. Azarbarzin and Z. M. K. Moussavi. Automatic and unsupervised snore sound extraction from respiratory sound signals. *IEEE Transactions on Biomedical Engineering*, 58(5):1156–1162, 2011. ISSN 00189294. doi: 10.1109/TBME.2010.2061846.

[9] I. R. Titze B. H. Story and E. A. Hoffman. The relationship of vocal tract shape to three voice qualities. *The Journal of the Acoustical Society of America*, 109(4):1651–1667, 2001.

[10] C. F. Babbs. Biomechanics of Snoring and Sleep Apnea. *Weldon School of Biomedical Engineering Faculty Working Papers*, Paper 9:1–114, 2017.

[11] T. S. Balint and A. D. Lucey. Instability of a cantilevered flexible plate in viscous channel flow. *Journal of Fluids and Structures*, 20(7 SPEC. ISS.):893–912, 2005. ISSN 08899746. doi: 10.1016/j.jfluidstructs.2005.05.005.

[12] R. Beck, M. Odeh, A. Oliven, and N. Gavriely. The acoustic properties of snores. *European Respiratory Journal*, 8(12):2120–2128, 1995. ISSN 09031936. doi: 10.1183/09031936.95.08122120.

[13] N. Ben-Israel, A. Tarasiuk, and Y. Zigel. Obstructive apnea hypopnea index estimation by analysis of nocturnal snoring signals in adults. *Sleep*, 35(9):1299–1305, 2012.

[14] C. D. Bertram. Flow-induced oscillation of collapsed tubes and airway structures. *Respiratory physiology & neurobiology*, 163(1-3):256–265, 2008.

[15] J. Blauert and N. Xiang. *Acoustics for Engineers*, chapter Horns and Stepped Ducts, pages 103–115. Springer-Verlag Berlin Heidelberg, 2009. doi: 10.1007/978-3-642-03393-3_8.

[16] J. Blauert and N. Xiang. *Acoustics for Engineers*, chapter Spherical Sound Sources and Line Arrays, pages 117–131. Springer-Verlag Berlin Heidelberg, 2009. doi: 10.1007/978-3-642-03393-3_9.

[17] J. Blauert and N. Xiang. *Acoustics for Engineers*, chapter Electromechanic and Electroacoustic Analogies, pages 27–35. Springer-Verlag Berlin Heidelberg, 2009. doi: 10.1007/978-3-642-03393-3_3.

[18] A. C. Den Brinker and O. Ouweltjes. SilentNight : algorithms for snoring characterization. Technical Report PR-TN-2017/00128, Philips Research Europe, 2017. Confidential until 12/2021.

[19] A. C. Den Brinker, M. Ergin, X. Aubert, and O. Ouweltjes. Measures of snoring and efficacy of anti-snoring devices. Technical Note PR-TN-2018/00073, Philips Research Europe, 2018. Confidential until 12/2022.

[20] P. E. Brockmann, F. Damiani, and D. Gozal. Sleep-disordered breathing in adolescents and younger adults: A representative population-based survey in chile. *Chest*, 149:981–990, April 2016. ISSN 1931-3543. doi: 10.1378/chest.15-2112.

[21] M. Brookes et al. Voicebox: Speech processing toolbox for matlab. *Software, available [Mar. 2011] from www. ee. ic. ac. uk/hp/staff/dmb/voicebox/voicebox. html*, 47, 1997.

[22] W. L. Capper, R. W. Guelke, and A. E. Bunn. The estimation of tube wall compliance using acoustic input impedance. *IEEE Transactions on Biomedical Engineering*, 38(6): 544–550, June 1991. ISSN 0018-9294. doi: 10.1109/10.81579.

[23] M. Cavusoglu, M. Kamasak, O. Erogul, T. Ciloglu, and Y. Serinagaoglu. An efficient method for snore / nonsnore classification. *Physiological Measurement*, 28:1–13, 2007. doi: 10.1088/0967-3334/28/8/007.

[24] D. Chalker and D. Mackerras. Models for representing the acoustic radiation impedance of the mouth. *IEEE transactions on acoustics, speech, and signal processing*, 33:1606–1609, 1985. ISSN 0096-3518. doi: 10.1109/tassp.1985.1164744.

[25] A. De Cheveigné and H. Kawahara. Yin, a fundamental frequency estimator for speech and music. *The Journal of the Acoustical Society of America*, 111(4):1917–1930, 2002.

[26] F. Chouly, A. Van Hirtum, P. Y. Lagrée, X. Pelorson, and Y. Payan. Numerical and experimental study of expiratory flow in the case of major upper airway obstructions with fluid-structure interaction. *Journal of Fluids and Structures*, 24(2):250–269, 2008. ISSN 08899746. doi: 10.1016/j.jfluidstructs.2007.08.004.

[27] L. Chuang, S. Lin S. Hsu, W. Ko, N. Chen, and Y. Tsai. Prevalence of snoring and witnessed apnea in taiwanese adults. *Chang Gung medical journal*, 31:175–181, 2008. ISSN 2072-0939.

[28] G. J. Dae, Y. C. Hae, R. R. Grunstein, and B. Yee. Predictive value of Kushida Index and acoustic pharyngometry for the evaluation of upper airway in subjects with or without obstructive sleep apnea. *Journal of Korean Medical Science*, 19(5):662–667, 2004. ISSN 10118934. doi: 10.3346/jkms.2004.19.5.662.

[29] F. Dalmasso and R. Prota. Snoring: Analysis, measurement, clinical implications and applications. *European Respiratory Journal*, 9(1):146–159, 1996. ISSN 09031936. doi: 10.1183/09031936.96.09010146.

[30] C. Doukas, T. Petsatodis, C. Boukis, and I. Maglogiannis. Automated sleep breath disorders detection utilizing patient sound analysis. *Biomedical Signal Processing and Control*, 7(3):256–264, 2012. ISSN 17468094. doi: 10.1016/j.bspc.2012.03.002. URL http://dx.doi.org/10.1016/j.bspc.2012.03.002.

[31] T. Drugman, B. Bozkurt, and T. Dutoit. A comparative study of glottal source estimation techniques. *Computer Speech and Language*, 26(1):20–34, 2012. ISSN 08852308. doi: 10.1016/j.csl.2011.03.003. URL http://dx.doi.org/10.1016/j.csl.2011.03.003.

[32] W. D. Duckitt, S. K. Tuomi, and T. R. Niesler. Automatic detection, segmentation and assessment of snoring from ambient acoustic data. *Physiological Measurement*, 27(10): 1047–1056, 2006. ISSN 09673334. doi: 10.1088/0967-3334/27/10/010.

[33] A. Tarasiuk E. Dafna and Y. Zigel. OSA severity assessment based on sleep breathing analysis using ambient microphone. *Proceedings of the 2013 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS)*, pages 2044–2047, 2013. ISSN 1557170X. doi: 10.1109/EMBC.2013.6609933.

[34] C. G. M. Fant. *Acoustic Theory of Speech Production*. Description and analysis of contemporary standard Russian. Mouton, 1970. URL https://books.google.nl/books?id=L9cFkAEACAAJ.

[35] C. G. M. Fant. *Acoustic Theory of Speech Production: With Calculations based on X-Ray Studies of Russian Articulations*. Description and Analysis of Contemporary Standard Russian. De Gruyter, 2012. ISBN 9783110873429. URL https://books.google.nl/books?id=UY0iAAAAQBAJ.

[36] JA Fiz, J Abad, R Jane, M Riera, MA Mananas, P Caminal, D Rodenstein, and J Morera. Acoustic analysis of snoring sound in patients with simple snoring and obstructive sleep apnoea. *European Respiratory Journal*, 9(11):2365–2370, 1996. ISSN 0903-1936. URL http://erj.ersjournals.com/content/9/11/2365.

[37] D. Flanagan, M. Arvaneh, and A. Zaffaroni. Audio signal analysis in combination with noncontact bio-motion data to successfully monitor snoring. In *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*, pages 3763–3766. IEEE, 2014.

[38] J. L. Flanagan. *Speech Synthesis*, chapter 3, pages 21–76. Springer Science & Business Media, 3 edition, 1972. doi: 10.1007/978-3-662-01562-9_6.

[39] J. S. Garofolo. Darpa timit acoustic-phonetic speech database. *National Institute of Standards and Technology (NIST)*, 15:29–50, 1988.

[40] E. R. Geddes and L. W. Lee. *Audio Transducers*. GedLee, 2002. ISBN 9780972208505.

[41] A.H. Gray and J.D. Markel. Cosh measure for speech processing. *The Journal of the Acoustical Society of America*, 58(S1):S97–S97, 1975.

[42] A.H. Gray and J.D. Markel. Distance measures for speech processing. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 24(5):380–391, 1976.

[43] R. H. Habib, R. B. Chalker, B. Suki, and A. C. Jackson. Airway geometry and wall mechanical properties estimated from subglottal input impedance in humans. *Journal of Applied Physiology*, 77(1):441–451, 1994. doi: 10.1152/jappl.1994.77.1.441. URL https://doi.org/10.1152/jappl.1994.77.1.441. PMID: 7961268.

[44] P. Harper, S. S. Kraman, H. Pasterkamp, and G. R. Wodicka. An acoustic model of the respiratory tract. *IEEE Transactions on Biomedical Engineering*, 48(5):543–550, 2001. ISSN 0018-9294. doi: 10.1109/10.918593. URL http://www.ncbi.nlm.nih.gov/pubmed/11341528{%}5Cnhttp://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=918593.

[45] P. V. Harper, H. Pasterkamp, H. Kiyokawa, , and G. R. Wodicka. Modeling and measurement of flow effects on tracheal sounds. *IEEE Transactions on Biomedical Engineering*, 50(1):1–10, 2003. ISSN 00189294. doi: 10.1109/TBME.2002.807327.

[46] M.H. Hayes. *Statistical digital signal processing and modeling*. John Wiley & Sons, 1996. ISBN 9780471594314. URL https://books.google.nl/books?id=N_VSAAAAMAAJ.

[47] L. Huang, S. J. Quinn, P. D. M. Ellis, and J. E. F. Williams. Biomechanics of snoring. *Endeavour*, 19(3):96–100, 1995. ISSN 01609327. doi: 10.1016/0160-9327(95)97493-R.

[48] K. Ishizaka, J. French, and J. Flanagan. Direct determination of vocal tract wall impedance. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 23(4):370–373, Aug 1975. ISSN 0096-3518. doi: 10.1109/TASSP.1975.1162701.

[49] F. Itakura. Analysis synthesis telephony based on the maximum likelihood method. In *The 6th international congress on acoustics, 1968*, pages 280–292, 1968.

[50] M. Kamil, C. Teng, and S. Hassan. Snoring and breathing pauses during sleep in the malaysian population. *Respirology (Carlton, Vic.)*, 12:375–380, May 2007. ISSN 1323-7799. doi: 10.1111/j.1440-1843.2007.01030.x.

[51] C. O. Kara, M. Zencir, N. Ardiç B. Topuz, and B. Kocagözolu. The prevalence of snoring in adult population. *Kulak burun bogaz ihtisas dergisi : KBB = Journal of ear, nose, and throat*, 14:18–24, 2005. ISSN 1300-7475.

[52] A. S. Karunajeewa, U. R. Abeyratne, and C. Hukins. Multi-feature snore sound analysis in obstructive sleep apnea-hypopnea syndrome. *Physiological Measurement*, 32(1):83–97, 2011. ISSN 09673334. doi: 10.1088/0967-3334/32/1/006.

[53] E.J. Kezirian, W. Hohenhorst, and N. de Vries. Drug-induced sleep endoscopy: the VOTE classification. *European Archives of Oto-Rhino-Laryngology*, 268(8):1233–1236, 2011.

[54] L. Lee, Y. Lo, J. Yu, G. Lee, Y. Ni, N. Chen, T. Fang, C. Huang, W. Cheng, and H. Li. Snoring sounds predict obstruction sites and surgical response in patients with obstructive sleep apnea hypopnea syndrome. *Scientific reports*, 6, 2016.

[55] A. D. Lucey, J. Wang J.J. Armstrong M.S. Leigh A. Paduch J H.Walsh D.D. Sampson A.J.C. King, G.A. Tetlow, P.R. Eastwood, et al. Measurement, reconstruction, and flow-field computation of the human pharynx with application to sleep apnea. *IEEE Transactions on Biomedical Engineering*, 57(10):2535–2548, 2010.

[56] S. Maeda. A digital simulation method of the vocal-tract system. *Speech communication*, pages 199–229, 1982. ISSN 01676393. doi: 10.1016/0167-6393(82)90017-6. URL http://www.sciencedirect.com/science/article/pii/0167639382900176.

[57] I. Marshall, N.J. Maran, S. Martin, M.A. Jan, J.E. Rimmington, J.J.K. Best, G.B. Drummond, and N.J. Douglas. Acoustic reflectometry for airway measurements in man: implementation and validation. *Physiological measurement*, 14(2):157, 1993.

[58] P. Mermelstein. Determination of the vocal-tract shape from measured formant frequencies. *The Journal of the Acoustical Society of America*, 41(5):1283–1294, 1967. ISSN 00014966. doi: 10.1121/1.1910470.

[59] S. Miyazaki, Y. Itasaka, K. Ishikawa, and K. Togawa. Acoustic analysis of snoring and the site of airway obstruction in sleep related respiratory disorders. *Acta oto-laryngologica. Supplementum*, 537(March):47–51, 1998. ISSN 0365-5237. doi: 10.1080/00016489850182738. URL http://www.ncbi.nlm.nih.gov/pubmed/9870649.

[60] V. Mohsenin. Effects of gender on upper airway collapsibility and severity of obstructive sleep apnea. *Sleep Medicine*, 4(6):523–529, 2003. ISSN 13899457. doi: 10.1016/S1389-9457(03)00168-0.

[61] A. K. Ng and T. S. Koh. Analysis and modeling of snore source flow with its preliminary application to synthetic snore generation. *IEEE Transactions on Biomedical Engineering*, 57(3):552–560, March 2010. ISSN 0018-9294. doi: 10.1109/TBME.2009.2034139.

[62] A. K. Ng, T. S. Koh, E. Baey, and K. Puvanendran. Speech-like analysis of snore signals for the detection of obstructive sleep apnea. *ICBPE 2006 - Proceedings of the 2006 International Conference on Biomedical and Pharmaceutical Engineering*, pages 99–103, 2006. doi: 10.1109/ICBPE.2006.348562.

[63] A. K. Ng, T. S. Koh, E. Baey, T. H. Lee, and K. Puvanendran U. R. Abeyratne. Could formant frequencies of snore signals be an alternative means for the diagnosis of obstructive sleep apnea? *Sleep Medicine*, 9(8):894–898, 2008. ISSN 13899457. doi: 10.1016/j.sleep.2007.07.010. URL http://dx.doi.org/10.1016/j.sleep.2007.07.010.

[64] A. K. Ng, E. Baey T. S. Koh and, T. H. Lee, U. R. Abeyratne, and K. Puvanendran. Role of upper airway dimensions in snore production: Acoustical and perceptual findings. *Annals of Biomedical Engineering*, 37(9):1807–1817, 2009. ISSN 00906964. doi: 10.1007/s10439-009-9745-7.

[65] 2005-2018 American Society of Clinical Oncology (ASCO).All rights reserved worldwide. Nasopharyngeal print, 2018. URL https://www.cancer.net/sites/cancer.net/files/nasopharyngealprint.jpg. [Online; accessed July 7, 2013].

[66] F. J. Owens and P.A. Lynn. *Signal Processing of Speech*. Palgrave Macmillan, 1993.

[67] J Rogelio Perez-Padilla, E Slawinski, LM Difrancesco, RR Feige, JE Remmers, and WA Whitelaw. Characteristics of the snoring noise in patients with and without occlusive sleep apnea. *American Review of Respiratory Disease*, 147(3):635–644, 1993.

[68] D. Pevernagie, R. M. Aarts, and M. De Meyer. The acoustics of snoring. *Sleep Medicine Reviews*, 14(2):131–144, 2010. ISSN 10870792. doi: 10.1016/j.smrv.2009.06.002. URL http://dx.doi.org/10.1016/j.smrv.2009.06.002.

[69] K. Qian, C. Janott, V. Pandit, Z. Zhang, C. Heiser, W. Hohenhorst, M. Herzog, W. Hemmert, and B. Schuller. Classification of the excitation location of snore sounds in the upper airway by acoustic multifeature analysis. *IEEE Transactions on Biomedical Engineering*, 64(8):1731–1741, 2017. ISSN 15582531. doi: 10.1109/TBME.2016.2619675.

[70] S. J. Quinn, L. Huang, P. D. Ellis, and J. E. Williams. The differentiation of snoring mechanisms using sound analysis. *Clinical Otolaryngology & Allied Sciences*, 21(2):119–23, 1996. ISSN 0307-7772. URL http://www.ncbi.nlm.nih.gov/pubmed/8735394.

[71] L. R. Rabiner and B. Juang. *Fundamentals of speech recognition*, volume 14, chapter 4.5. PTR Prentice Hall Englewood Cliffs, 1993. ISBN 0130151572.

[72] M. P. Robb, J. Yates, and E. J. Morgan. Vocal tract resonance characteristics of adults with obstructive sleep apnea. *Acta Oto-Laryngologica*, 117(5):760–763, 1997. ISSN 00016489. doi: 10.3109/00016489709113474.

[73] S. Saha, T. D. Bradley, M. Taheri, Z. Moussavi, and A. Yadollahi. A subject-specific acoustic model of the upper airway for snoring sounds generation. *Scientific Reports*, 6:1–10, 2016. ISSN 20452322. doi: 10.1038/srep25730. URL http://dx.doi.org/10.1038/srep25730.

[74] H. Schlichting and K. Gersten. *Boundary-Layer Theory*, chapter LaminarTurbulent Transition, pages 415–496. Springer-Verlag Berlin Heidelberg, Heidelberg, 2017.

[75] B. Schuller, S. Steidl, A. Batliner, E. Bergelson, J. Krajewski, C. Janott, A. Amatuni, M. Casillas, A. Seidl, M. Soderstrom, A. S. Warlaumont, G. Hidalgo, S. Schnieder, C. Heiser, W. Hohenhorst, M. Herzog, M. Schmitt, K. Qian, Y. Zhang, G. Trigeorgis, P. Tzirakis, and S. Zafeiriou. The INTERSPEECH 2017 computational paralinguistics challenge: Addressee, cold & snoring. In *Proceedings of Interspeech 2017*, pages 3442–3446, Stockholm, Sweden, 2017. doi: 10.21437/Interspeech.2017-43.

[76] B. Suki, R. H. Habib, and A. C. Jackson. Wave propagation, input impedance, and wall mechanics of the calf trachea from 16 to 1,600 hz. *Journal of Applied Physiology*, 75(6): 2755–2766, 1993. doi: 10.1152/jappl.1993.75.6.2755. URL https://doi.org/10.1152/jappl.1993.75.6.2755. PMID: 8125900.

[77] D. Teculescu, B. Hannhart, A. Cornette, B. Montaut-Verient, J.M. Virion, and J.P. Michaely. Prevalence of habitual snoring in a sample of french males. *Respiration*, 68 (4):365–370, 2001.

[78] H. Wakita. Direct estimation of the vocal tract shape by inverse filtering of acoustic speech waveforms. *IEEE Transactions on Audio and Electroacoustics*, 21(5):417–427, 1973.

[79] H. Wakita and C. G. M. Fant. Toward a better vocal tract model. *STL-QPSR*, 19:009–029, 1978.

[80] S. O. Wali and B. A. Abaalkhail. Prevalence and predictors of habitual snoring in a sample of saudi middle-aged adults. *Saudi Medical Journal*, 36:920–927, August 2015. ISSN 0379-5284. doi: 10.15537/smj.2015.8.11848.

[81] B. L. Welch. The generalization of 'student's' problem when several different population variances are involved. *Biometrika*, 34(1/2):28, 1947. ISSN 0006-3444. doi: 10.2307/2332510.

[82] G. R. Wodicka, K. N. Stevens, H. L. Golub, E. G. Cravalho, and D. C. Shannon. A model of acoustic transmission in the respiratory system. *IEEE Transactions on Biomedical Engineering*, 36(9):925–934, 1989.

[83] H. Xu, W. Huang, L. Yu, and L. Chen. Sound spectral analysis of snoring sound and site of obstruction in obstructive sleep apnea syndrome. *Acta Oto-Laryngologica*, 130(10): 1175–1179, 2010. ISSN 00016489. doi: 10.3109/00016481003694774.

[84] Azadeh Yadollahi and Zahra Moussavi. Automatic breath and snore sounds classification from tracheal and ambient sounds recordings. *Medical Engineering Physics*, 32(9):985 – 990, 2010. ISSN 1350-4533. doi: https://doi.org/10.1016/j.medengphy.2010.06.013. URL http://www.sciencedirect.com/science/article/pii/S1350453310001402.

[85] M. Zhao, T. Barber, P. A. Cistulli, K. Sutherland, and G. Rosengarten. Simulation of upper airway occlusion without and with mandibular advancement in obstructive sleep apnea using fluid-structure interaction. *Journal of Biomechanics*, 46(15):2586 – 2592, 2013. ISSN 0021-9290. doi: https://doi.org/10.1016/j.jbiomech.2013.08.010. URL http://www.sciencedirect.com/science/article/pii/S0021929013003904.