Physical Modelling of the Vocal Tract with the 2D Digital Waveguide Mesh

Jack Mullen

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy to the
Department of Electronics

THE UNIVERSITY of York

April 2006

Abstract

Acoustical physical modelling synthesis uses mathematical algorithms to describe a real-world sound production process or propagational environment. Digital waveguides can be used to form a 1D model of the vocal tract, simplistically represented as a series of cylindrical tubes of varying radius along a straight axis. This 1D signal propagating element can also be extended to create a digital waveguide mesh (DWM), giving acoustical synthesis of a higher dimensional structure, such as a 2D surface or 3D space.

The work contained in this thesis is an investigation into the effects of increased dimensionality in the 1D waveguide vocal tract paradigm. A 2D DWM is configured as a model of the tract, such that shape characteristics are set within the width of the mesh. Wave propagation and reflection is simulated along the tract from the glottis to the lips, as well as across it, between the two inner walls, thereby removing plane-wave limitations inherent in the 1D model. The 2D tract is found to give accurate formant synthesis, producing vowels that give a good match to real-world targets. However, problems associated with high sampling frequency limitations and discontinuous dynamic operation are identified. Movements readily occurring in speech, such as diphthongs, are not easily accommodated by the static mesh structure.

A novel alternative approach is also presented which maintains a rectangular mesh, but maps the changing tract shapes onto the waveguide impedances. This allows for stable dynamic manipulation of the modelled space. Furthermore, sampling frequency limitations are removed, such that real-time operation and interaction with the 2D tract model is achieved.

Contents

List of Figures

Acknowledgments

Firstly, I would like to express my sincere gratitude to my supervisors in the department of Electronics at the University of York; Professor David Howard and Dr. Damian Murphy. From the outset, their continuous support, guidance and enthusiasm gave me a strong and reliable infrastructure in which to conduct the research contained within this thesis. In many discussions on the human voice, David's broad knowledge and ability to inspire were fundamental in building my own understanding, and in giving momentum and direction to the work. Damian's extensive expertise in physical modelling, and the generosity with which he shared it with me, were significant factors in my own motivation and accomplishments. To both, I am indebted.

I would also like to mention those who have inspired and enlightened me, intentionally or unintentionally, throughout the course of my research. Thanks to Matti Karjalainen and Paavo Alku from the Helsinki University of Technology, and Christer Gobl from Trinity College Dublin, who all helped me to resolve issues on glottal excitation in conversations at conferences and communication via email. Further thanks go to Dr. John Szymanski at the department of Electronics at York, who has a wonderful ability of explaining complicated mathematical concepts in understandable terms. I am also grateful of the supportive network of friends and associates that exists within the department. My appreciation goes to Mark Beeson, Rod Fry, Carl Hetherington, John Matthews, and everybody in the Audio Lab, who have advised me on programming, signal processing and audio matters.

Finally, I would like to thank my parents, Chris and Oriole Mullen for their love and support during all my time in York.

Declaration

I hereby declare that this thesis is entirely my own work and all contributions from outside sources, through direct contact or publications, have been explicitly stated and referenced. I also declare that some parts of this program of research have been presented previously, at conferences and in journals. These publications are listed as follows:

- **Real-Time Dynamic Articulations in the 2D Waveguide Mesh Vocal Tract Model**, J. Mullen, D. H. Howard and D. T. Murphy, journal paper accepted on 17th February 2006 to *IEEE Transactions on Audio, Speech and Language Processing*, awaiting publication [1].
- **Waveguide Physical Modelling of Vocal Tract Acoustics: Flexible Formant Bandwidth Control from Increased Model Dimensionality**, J. Mullen, D. H. Howard and D. T. Murphy, journal paper accepted on 25th April 2005 to *IEEE Transactions on Speech and Audio Processing*, to be published in May 2006 [2].
- **Speech Synthesis Using Multidimensional Physical Modelling**, poster presentation at *Set For Britain: Annual Presentations by Britain's Top Younger Scientists, Engineers and Technologists*, at the Houses of Commons, London, March 2005.
- **Acoustical Simulations of the Human Vocal Tract Using the 1D and 2D Digital Waveguide Software Model**, J. Mullen, D. H. Howard and D. T. Murphy, paper and poster presentation at *7th International Conference on Digital Audio Effects (DAFx-04), Naples, Italy* [3].
- **Digital Waveguide Mesh Modelling of the Vocal Tract Acoustics**, J. Mullen, D. H. Howard and D. T. Murphy, paper and oral presentation at *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA-03), New York, USA* [4].

Jack Mullen
April 2006

Chapter 1

Introduction

1.1 Speech Synthesis

Current technologies in artificial speech generation are at a stage of near perceived realism. Many state of the art text-to-speech (TTS) systems use sample-based concatenative synthesis [5], [6]. Instances of the fundamental units of speech, phonemes, are taken from recordings of natural speech and spliced together to create new words and sentences not present in the original utterance. The recordings are of a spoken voice, typically that of an actor, reading aloud for several hours. A large database is constructed as each possible diphone (the transition between the middle of one phoneme and another) is extracted and categorised to be recalled for concatenation at a later time. Such a scheme is, however, restricted to regeneration only of sounds present in the original recordings. Even limited processing of the signal may prove detrimental to the resulting naturalness. Furthermore, the user may only communicate with the vocal identity provided with the system.

Articulatory vocal tract modelling attempts to recreate the behaviour of the human speech apparatus to simulate the process of speaking, rather than simply the sounds it generates. The tract is an acoustic resonator with various mouth features that constantly alter, constrict and stop the way vibrations travel through it. These articulations and tract shape changes, combined with the glottal source produce the sounds we perceive as speech.

An effective vocal tract model will demonstrate the ability to dynamically adapt, accommodating the changes in the tract area function. This allows for production of a connected chain of speech sounds to form continuous utterances. Such a dynamic model can be used to simulate a diphthong - a slide between two vowels, for example /aʊ/ in the word *house*. Area function changes can also be made to represent constrictions to the air flow in the model, giving lateral articulation such as the /l/ in the word *lip*. Plosive articulations can be modeled in a similar way, forcing a momentary stop and then release of the air flow, such as the /p/ in the word *put*. Articulatory vocal tract simulations are physical models, in that they approximate observations made in natural speech production.

1.2 Physical Modelling

Discretisation of continuous variables in a real-world system for computer-based synthesis is a common modelling technique. Acoustical physical modelling is currently a very interesting subcategory within this field. Well-established finite difference techniques have been developed for analysis and simulation of field variables in electromagnetics [7] and fluid dynamics [8]. This theory and methodology can be equated to problems in sound and vibrations, and is continually adapted for applications in the audio spectrum [9].

The formal definition of the 1D digital waveguide as an acoustical physical model can be accredited to Julius Orion Smith III from the Center for Computer Research in Music and Acoustics (CCRMA) in Stanford University, USA. His online book is an extensive source for reference [10]. The waveguide is typically used as the fundamental building block that facilitates acoustic signal propagation in distributed media. Much of the following research has been directed towards its use in 1D virtual musical instruments, for example forming the acoustic bore along the length of a clarinet [11]. Similarly, its extension into higher dimensional representations to form a 2D or 3D digital waveguide mesh (DWM) has been documented

[12] [13]. In 2D, the DWM can be used to form a model of a vibrating plate or surface, such as a drum skin [14]. It can also be used as a efficient precursor to a 3D model for examining wave propagation and testing new techniques whilst exploiting the lower computational demands inherent in reduced dimensionality. A 3D DWM model can be constructed that provides a virtual simulation of the acoustics of a room [15]. The waveguide mesh naturally includes the diffraction effects that are absent in alternative acoustical simulation methods, such as image-source [16] and ray-tracing [17]. It therefore provides an accurate, although often computationally intensive, model of sound propagation within an enclosed structure.

1.3 Physical Modelling of the Vocal Tract

The Kelly-Lochbaum piecewise articulatory vocal tract analogy is commonly cited as the archetypal physical model [18]. One-dimensional wave propagation along the tract is simulated with a chain of waveguides. The spatial sampling used implies that the model represents a series of adjoining cylinders of varying radius along a straight axis. The system acts as a filter, shaping the spectrum with what is known in vocal terms as *formants*, to the glottal input signal, such that speech-like sound is observed at the output. Such time-domain articulatory vocal tract models have been used to synthesise simple words with near realism, but the method is not comparable to concatenative synthesis at its current state of development.

Several studies have been conducted into a 3D frequency-domain vocal tract model using finite element modelling (FEM) [19] [20]. Limited research has been used to demonstrate a time-domain model using a 2D transmission line matrix (TLM) [21], which is equivalent to the DWM. Results have shown that cross-tract modal patterns are present in the output.

The work contained within this thesis is a continuation of the ideas associated with the TLM tract representation. Here it is presented as an investigation into the effects of increased dimensionality in the time-domain Kelly-Lochbaum vocal tract model. It is proposed that a 2D DWM model of

the air-cavity contained within the human speech system may be constructed. The additional dimensionality includes propagational effects across the tract as well as along it, and so should offer increased accuracy of synthesis. Indications are made towards how the techniques discussed here could be applied to an eventual full 3D model, which should present a highly accurate articulatory physical model of the voice.

1.4 Motivation

Concatenative synthesis is currently the preferred method of artificial speech generation, giving the most natural results. However, as the requirements of a system increase, the database needed to contain all the recorded samples grows. Extended demands of such a system might involve elements of speech across different languages, or the inclusion of emotional expression. Clearly, such a speech synthesis system offering extensive naturalness would imply an enormous database of samples. This opinion is echoed in the 2001 book *Multilingual Text-To-Speech Synthesis - The Bell Labs Approach*

> *"... we feel that the complex forms of coarticulation found in human speech can only be mimicked by accurate articulatory models, because concatenative synthesis would require too many units to achieve these - significantly higher - levels of quality."* [22]

A well-established time-domain articulatory vocal tract model is the piecewise cylinder analogy proposed by Kelly and Lochbaum [18]. Many simplifications are made in forming it as a 1D representation. The simplifying assumption that is removed in this work is the limitation of the propagational mechanism to longitudinal plane wave motion. This implies an extension of the dimensional representation of the tract air-cavity within the model. It has been indicated in the surrounding literature that this may be worth consideration. In Perry Cook's 1996 summary of singing voice synthesis he suggests that future work in speech physical models might involve

> *"...some significant component of non-linearity, and/or higher dimen-*

> *sional models. The main research areas involve modeling of airflow in the vocal tract, development of more exact models of the inner shape of the vocal tract tube, physical models of the tongue and other articulators, more accurate models of the vocal folds, and facial animation coupled to voice synthesis."* [23]

The eventual goal of a realistic articulatory synthesiser is still a distant ideal. Natural speech production involves many intricate and slight muscular movements in the lips, tongue and jaw, organised with carefully timed synchrony. A high-representation articulatory model that includes all of these features would clearly be a very complicated system with large computational demands. In terms of general speech research, each of the individual articulators and features themselves present interesting analysis and modelling challenges.

The vocal subsystem under scrutiny in this work is the propagational airway contained within the vocal tract, with a view to making fewer shape-based simplifications.

1.5 Thesis Outline

To begin with, physical modelling theory is discussed at length in Chapter 2. Derivations are given for the fundamental mathematical and physics-based concepts which are used in discrete-time simulations of continuous variable systems. In particular, the digital waveguide is presented as a physical model which provides 1D propagation based on the travelling wave solution to the wave equation. The review then moves on to consider how such a 1D representation can be extended to form a model of a higher dimensional structure with a DWM. Relevant additional mathematical material is included in Appendix A.

Chapter 3 gives a selective and brief overview of the vast field of research that is the human voice. The IPA phonetic notation is used to identify speech sounds throughout the thesis. Details on this and example words are given in Appendix B. The vocal tract is introduced as an acoustic resonator which

imparts spectral formant characteristics onto the glottal excitation. Details follow on some of the aspects of speech, such as production of vowels and diphthongs, and plosive consonants. Various methods of generating artificial speech are outlined. In particular, attention is drawn to the 1D time-domain waveguide vocal tract analogy, and to attempts that have been made to improve the model.

The construction of the 2D DWM vocal tract model is examined in Chapter 4. Two approaches are identified for mapping the vocal tract area function onto the mesh width; one based on a radial representation, and one based on a diametral 2D slice though a 3D tract. The diametral method is selected for further analysis on the *widthwise* mapped model, because of its ease of implementation and the intuitive way in which it leads to a full 3D model. Two interpretations for the diametral approach are used in simulations; the width can be set as proportional to the tract radius, or to the radius-squared. The model is shown to produce system frequency spectra with formants that are similar to those generated with an equivalent 1D model. It is also noted that the extent to which these formants match, and the resulting likeness to the modelled vowel, is improved when the effect of the shape changes is enhanced with the radius^2 area functions. Next, the sensitive and linear nature of formant bandwidth variation is demonstrated when the additional boundary reflection coefficient is used to regulate energy losses in the system. Some of the disadvantages of the width mapped mesh tract model are highlighted with an analysis of the dynamic capabilities of the system.

Limitations associated with the widthwise mapped mesh are addressed in the design of a novel alternative approach presented in Chapter 5. The area function is applied to the waveguide impedances, rather than mesh width, such that it retains its structure whilst its resonant properties are altered. Two approaches are used for mapping the area function onto the model; impedance values can be defined as proportional to the radius-squared, or to the radius-cubed. This method is shown to allow for dynamic shape changes to be made, and for a reduction in sampling frequency, leading to

construction of real-time interactive software to demonstrate the model.

Chapter 6 presents a discussion on the results obtained from both models and an exploration of directions for potential future work.

A general conclusion is presented in Chapter 7. Relevant materials from this research are available in the form of an accompanying CD affixed in the back of paper copies of this thesis, and from the internet at `http://www-users.york.ac.uk/~dtm3/vocaltract/`. The files include the software that has been developed to demonstrate the real-time dynamic impedance mapped DWM vocal tract model, and various sound examples created with both model types.

1.6 Specific Contributions of the Research

The novel contributions contained in this thesis are as follows.

- An analysis of a 2D DWM time-domain simulation of the vocal tract is presented.
- The flexible formant bandwidth response that is obtained with the additional waveguide mesh boundary in the width-mapped model is demonstrated.
- Waveform discontinuities are highlighted as a problematic issue in dynamic operation.
- A new method of changing the shape of the modelled space using impedance mapping is shown to allow stable dynamic changes to a DWM.
- This is the first demonstration of a stable, dynamically varying DWM and potentially opens new perspectives for research into vocal synthesis using multidimensional signal processing, and also in more general DWM applications.
- A real-time response is achieved in the 2D DWM vocal tract model.

The work indicates that there should be clear advantages in moving towards a full 3D model, giving extensive control over many physical parameters of the human vocal system.

Chapter 2

Physical Modelling Synthesis

2.1 Introduction

The main aim in a computer-based simulation of the sound produced by a musical instrument is naturalness of synthesis. In order to achieve this the system will be required to generate sound that it is perceived as being very similar to, or ultimately, indistinguishable from, the instrument in question. Well established methods based on the play-back of recorded samples [24], or reproduction of the known spectral content [25] have traditionally been used to create a digital instrument. However, problems that are inherent with such methods can be detrimental to the resulting naturalness. Sample-based methods only allow for reproduction of content that is present in the memory *wavetable*, and manipulation of the stored sounds often yields artificial results [26]. Controlling parameters used in spectral synthesis, such as frequency, magnitude and phase are mathematical concepts. They have little in common with terms associated with musical performance and therefore have limited interpretation for non-engineers [27].

A mathematical description of a real-world process or system can be used to produce a simplified virtual simulation of the bodies within the defined problem domain and the forces acting on them. This direct representation allows for interaction with the model in a manner not possible using other synthesis methods. Parameters controlling a system based on reconstruction

of the physical process itself are based on, and therefore explicitly linked to, those governing system response. Such intuitive, semantic control is inherent in a *physical model*. In addition, the principles behind physical modelling generally allow for low memory requirements to be placed upon a system as the underlying algorithms involved are minimal. The definition of a set of laws outlining possible modes of operation also allows for experimentation with actions not permissable in the real world, for example on the grounds of safety or impracticality.

With an increase in both the extent of knowledge about the physical workings of the real-world instrument and the available computing power with which to demonstrate it, it becomes feasible to increase the depth of representation in the simulation. A physical model recreates the sound generation process, rather than the simply the resultant sound. It therefore behaves in a manner that is similar to its real-world equivalent, producing sound of a highly organic nature [28].

Acoustical physical modelling uses the discretisation of real world mechanics to capture essential aspects of a sound producing or propagating system. For example, a vibrating body can be represented as a series of interconnected elements where each obeys the physics-derived laws governing interaction with its neighbours. With constraints applied to the system and inputs defined, the virtual model exhibits natural behaviour that approximates real world expectations. One of the earliest physical models was a discrete-time representation of the 1D acoustics of the vocal tract [18]. Following developments included calculations of the motion of a vibrating string [29] and discrete mass-and-spring oscillatory systems [30]. Figure 2.1 shows an example of how a series of interconnected masses and springs can be used to represent a string with fixed ends. An initial displacement is applied and the model exhibits oscillatory motion similar to a real world string.

The use of the *digital waveguide* as the signal propagating unit within a physical model was formally defined when the Karplus-Strong plucked string algorithm [31] was extended by Julius Orion Smith III into a more

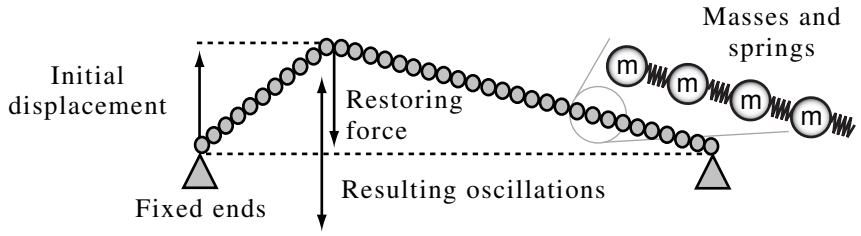


Figure 2.1: *A physical model of a string using the mass and spring paradigm*

general case [32]. In recent years digital waveguide modelling has emerged as popular and versatile method of acoustical synthesis. The extent to which the model follows behaviour stated in theory is the determining factor in giving accurate enough representation while maintaining suitability for given applications. Heavily simplified models might be employed where real-time response is of concern, for example in the 1D synthesis of a small resonating structure like a musical instrument [33] [34] [35] or a collision between two objects [36]. Vibrations in structures with higher dimensionality can be simulated, such as the 2D plane formed across a drum skin [14]. The use of waveguide modelling in systems with a focus on accuracy of synthesis would embrace more of the facets of the underlying theory used to define the model, such as the air absorption effects and diffuse reflections used in 3D room acoustics simulation [13] [37] [15].

This chapter examines the methods used to define a physical model of an acoustic system, beginning with the underlying theory. The physical constants and phenomenon that occur in the definitions of sound and vibrations are discussed. These parameters are used in the derivation of continuous equations that describe oscillatory wave motion. Some of the techniques used to make the model suitable for computer simulation are discussed. Emphasis is placed on the process of solving the wave equation at points within a given discretisation. One such technique, the digital waveguide physical model, is examined in depth. In particular, it is shown how this representation of 1D wave propagation can be extended to form a *digital waveguide mesh* that is used to model structures with higher dimensionality.

2.2 Acoustical Theory

The definition of the word *sound* follows two forms. It describes both the cause and effect in the process of hearing. In perceptual terms

> *"Sound is the auditory sensation produced through the ear by the alteration in pressure, particle displacement, or particle velocity propagated in an elastic material"* [38]

In terms of the physical process which results in perception, sound is also the propagation of the mechanical disturbances of the particles within the medium. It is this vibrational form of sound that is analysed and numerically simulated in the field of acoustical physical modelling synthesis.

2.2.1 The Velocity of Sound in Air

A sound wave propagating in a gas exhibits longitudinal motion. It travels away from a source as a series of compressions and rarefactions in the particles. A momentary localised density increase gives rise to nearby decreases. A *pressure* gradient is formed across the change in density and elastic forces in the particles attract them back towards their rest position. This particle movement implies a *velocity* component in the resulting wave. The restoring displacement back towards the source creates further compressions with neighbouring particles, and hence further rarefactions, which are also restored by elastic forces. This process continues as both the pressure and velocity components of the wave propagate away from the source with the same speed, maintaining a phase difference of $\pi/2$ between them. The total velocity at which an arbitrary wave moves away from the source is referred to as the wave speed v. It is determined from first principles using the ideal gas law and the universal gas constant $R = N_A k$, with Avagadro's number $N_A = 6.022 \times 10^{23}$ mol^{-1} and Boltzmann's constant $k = 1.3807 \times 10^{-23}$ JK^{-1}. In general, the ambient pressure P of a gas of density ρ, molecular mass m and absolute temperature in Kelvin T_K, is [39]

$$P = \frac{\rho}{m} R T_K \tag{2.1}$$

The velocity v of a wave in a gas with heat capacity ratio γ is given by

$$v = \sqrt{\frac{\gamma P}{\rho}} \tag{2.2}$$

Substitution of (2.1) into (2.2) shows that the wave speed in a gas is independent of pressure [40]

$$v = \sqrt{\frac{\gamma R T_K}{m}} \tag{2.3}$$

If the gas is air, then the heat capacity ratio is $\gamma = 7/5$, molecular mass is $m = 29.0 \times 10^{-3}$ kg mol^{-1} and the gas constant is $R = 8.314$ J mol^{-1} K^{-1}. The speed of a sound wave in air is usually denoted by the character c, and at temperature T_K is calculated as

$$c = 20.03\sqrt{T_K} \tag{2.4}$$

An equivalent form of (2.4) uses temperature in Celsius $T_C = T_K - 273$

$$c = 20.03\sqrt{273 + T_C} \tag{2.5}$$

A linear approximation for the temperature variation determines the speed of sound in air at temperature T_C as [41]

$$c = 331 + 0.6T_C \tag{2.6}$$

As such, a sound wave travels through air at room temperature of $T_C = 20°$C at an approximate speed of $c = 343$ ms^{-1}.

2.2.2 Acoustic Impedance and Admittance

The two components of the wave - the scalar pressure variation p, and the velocity vector of the particle movements $\underline{u}$ - are closely interlinked. They follow the relationship that their ratio in an unbounded homogenous medium is always constant. This quantity is called acoustic impedance Z.

In some cases it is convenient to refer to it in terms of its inverse, acoustic admittance Y.

$$Z = \frac{1}{Y} = \frac{p}{u} \tag{2.7}$$

Impedance is defined as the geometric mean of the two sources of resistance to displacement in the medium: tension and mass [10]. It can be calculated from the density ρ and Young's modulus of elasticity E of the medium to be

$$Z = \sqrt{\rho E} \tag{2.8}$$

The Young's modulus of a gas is defined as $E = \gamma P$. In the context of air, the wave speed as calculated from (2.2), becomes

$$c = \sqrt{\frac{E_{air}}{\rho_{air}}} \tag{2.9}$$

Rearranging this for E and substituting into (2.8) gives

$$Z_{air} = \rho_{air} c \tag{2.10}$$

Air at a temperature of $T_C = 20°\text{C}$ has density of approximately $\rho_{air} = 1.2$ kg m^{-3}. Therefore the acoustic impedance of air at room temperature is $Z_{air} = 411.6$ kg m^{-2} s^{-1}. Similarly, the acoustic admittance of air is $Y_{air} = 2.43 \times 10^{-3}$ kg^{-1} m^2 s.

It is also of interest to define the characteristic impedance experienced by a wave propagating through an enclosed air column. For example, a wave travelling through an infinitely long tube experiences an impedance that is equivalent to Z_{air} spread across the cross sectional area. The characteristic impedance Z and admittance Y of an acoustic tube of cross-sectional area A is

$$Z_{tube} = \frac{1}{Y_{tube}} = \frac{\rho c}{A_{tube}} \tag{2.11}$$

Many parallels may be drawn on the relationship between acoustic impedance, pressure and velocity outlined in this section, and the relation-

ship between electrical impedance, voltage and current in circuit theory [9].

2.3 The Wave Equation

The wave equation describes propagation of the particle disturbances within a given coordinate system and speed. For a system in which variables of interest are considered as a function of time only, an ordinary differential equation (ODE) is constructed to describe the oscillatory behaviour. In the more complex case where dependent variables change with time as well as one or more spatial coordinates, then a partial differential equation (PDE) is used.

2.3.1 Simple Harmonic Motion

When an ideal mass and spring are connected together they form a lossless second order resonator. A more realistic, lossy system can be created with the introduction of a damper. A simple damped resonator is illustrated in Figure 2.2, where a small displacement results in oscillations created by the spring restoring force. The damped harmonic motion will eventually decay to return the mass to its equilibrium position.

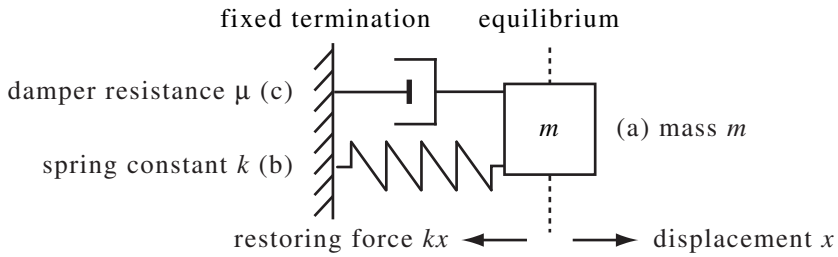


Figure 2.2: *Components in a simple resonator: (a) a mass, (b) spring and (c) damper*

Simple harmonic oscillations in the system are defined by identifying the three forces acting on the mass making the following assumptions:

- Mass (a) moves only a small distance in the x-axis and is unaffected by any external force such as gravity or friction. The net force on the mass

is defined by Newton's Second Law as:

$$f_n = m\frac{d^2x}{dt^2} \tag{2.12}$$

- Component (b) has spring constant k, and is defined as massless with no damping factor. Hooke's Law states that the spring restoring force against the displacement is:

$$f_r = -kx \tag{2.13}$$

- The force required to overcome the damper (c) is typically approximated as proportional to the velocity of oscillations. It is also assumed as massless. If μ is the damper resistance, the force against displacement is:

$$f_d = -\mu\frac{dx}{dt} \tag{2.14}$$

Newtonian Mechanics requires that $f_n = f_r + f_d$. The system equation is therefore a second order ODE.

$$m\frac{d^2x}{dt^2} + \mu\frac{dx}{dt} + kx = 0 \tag{2.15}$$

If the damper was to be disconnected, or $\mu = 0$, then the natural frequency of undamped oscillations is $\omega_0 = \sqrt{k/m}$. With the influence of the damper a decay constant emerges as $\alpha = \mu/(2m)$, and the damped system oscillations occur at a lower frequency of $\omega_d = \sqrt{\omega_0^2 - \alpha^2}$.

2.3.2 One-Dimensional Wave Motion

Oscillatory wave motion in 1D can be defined by considering the forces acting on an ideal elastic body under tension [42]. Figure 2.3 illustrates a short string of density ρ that lies at rest along the horizontal x-axis and is given a small initial displacement in the vertical y-axis, resulting in a tension T. The assumption that negligible movement of the string takes place in the horizontal direction allows for the Newtonian derivation of acceleration in

terms of the acting force using only the transverse component of the tension. The description of wave motion is therefore simplified to 1D. Additionally, this derivation implies that propagation on the string is non-dispersive (all frequencies travel with the same speed) and lossless in both directions.

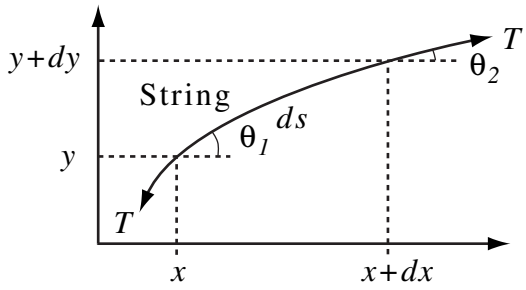


Figure 2.3: *String displacement resulting in constant tension T*

The string and amplitude of displacement are both small enough such that $ds \approx dx$. The Tension T acts as $T\sin(\theta_1)$ at x, and $T\sin(\theta_2)$ at $x+dx$, such that the total force acting on the string is $T\sin(\theta_2) - T\sin(\theta_1)$. From Newton's Second Law (2.12), then

$$T[sin(\theta_2) - \sin(\theta_1)] = \rho dx \frac{\partial^2 y}{\partial t^2} \tag{2.16}$$

If θ is small enough then $\sin(\theta) \approx \tan(\theta) = \frac{\partial y}{\partial x}$, it is possible to say

$$\frac{T}{dx}\left[\left.\frac{\partial y}{\partial x}\right|_{x+dx} - \left.\frac{\partial y}{\partial x}\right|_{x}\right] = \rho \frac{\partial^2 y}{\partial t^2} \tag{2.17}$$

This difference in gradients is the differential coefficient of the gradient $\frac{\partial y}{\partial x}$ divided by the space interval dx. Hence, as dx tends towards zero this becomes a second order PDE

$$\frac{\partial^2 y}{\partial x^2} = \frac{\rho}{T}\frac{\partial^2 y}{\partial t^2} \tag{2.18}$$

This can be rearranged to obtain the classic form of the one dimensional wave equation. It describes the acceleration of the wave in the y direction in terms of the second derivative of its resulting displacement with respect to x, at a

wave speed of $v = \sqrt{T/\rho}$

$$\frac{1}{v^2}\frac{\partial^2 y}{\partial t^2} = \frac{\partial^2 y}{\partial x^2} \tag{2.19}$$

In the context of a 1D acoustic pressure wave p travelling along an ideal uniform acoustic tube in the x direction at the speed of sound c, the PDE becomes

$$\frac{1}{c^2}\frac{\partial^2 p(x,t)}{\partial t^2} = \frac{\partial^2 p(x,t)}{\partial x^2} \tag{2.20}$$

2.3.3 Boundary Conditions and Resonance

When the propagating medium under scrutiny is bounded, this implies that *reflections* will take place. The manner in which these reflections manifest and the physical properties of the system collectively define how oscillations begin to appear within the system. At certain frequencies, oscillations exhibit increased amplitude where a greater transfer of energy exists. Standing waves are formed with wavelengths that are directly related to the geometry of the space defined within the bounds. This phenomenon is called *resonance*.

Boundary conditions can be divided into two types [43]. Firstly, *rigid* surfaces are defined such that the bounding impedance is very much bigger that that of the transmitting medium. For a string this implies that one end is fixed and no movement takes place in the y direction. In other words it is a displacement *node*. When a wave is incident upon the fixed end of the string, it exerts a transverse force on the termination. As no movement takes place, Newton's Third Law dictates that the rigid boundary must be exerting an equal and opposite force on the string. The reflection at the fixed end of a sting therefore presents a *phase inversion*.

Secondly, a *free-end* exists where the bounding impedance is very much lower than the transmitting medium. This can be thought of as a string that is attached to a frictionless guide-rod which prohibits motion in the x axis, but allows unhindered displacement in the y direction. Here, maximal movement, with reference to the amplitude of the wave, may take place. It is a displacement *antinode*. A longitudinal force is exerted on the string from the connection in the x direction, but not in the transverse y direction and so

energy is reflected with a *phase preservation*.

The phase properties of reflected longitudinal waves at free and rigid ends of an acoustic tube follow an inverse relationship to those of transverse waves on a string [44]. Modal analysis of a tube can be achieved with two wave variables. Here, pressure is the amplitude of variations above and below atmospheric pressure. Displacement refers to the distance away from rest position that the air molecules have moved in order to create this pressure change. A rigid boundary constitutes a closed end of a tube where no particle movement may take place. This implies a pressure antinode where the force exerted by the closed end is large enough to equal the maximal pressure presented by the wave, in order to enforce this displacement node. When a high pressure part of the longitudinal wave hits the boundary it exerts a force perpendicularly towards it. The boundary in return exerts a force in the opposite direction. The wave is reflected as a high pressure. There has, therefore, been a phase preserving reflection.

An open ended tube constitutes a free-end reflection. Atmospheric pressure level is present at this end and so the amplitude of pressure variations is equal to zero. This is a pressure node. Air molecules are free to move in the unbounded region and so a displacement antinode is formed. Some proportion of a high pressure part of an incident wave is transmitted through the open end. As this happens, a small amount of air is sucked out of the tube with it. Consequently, a low pressure region is reflected back into the tube. This forms a phase inverting reflection.

2.3.4 Wave motion in 3D Cartesian Space

Equation 2.20 is an example of where wave motion in 1D has been extrapolated from the more general 3D case using a separation of variables [43]. The 3D wave equation in cartesian coordinates x, y and z, is

$$\frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} = \frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} + \frac{\partial^2 p}{\partial z^2} \tag{2.21}$$

A brief review of different coordinate systems is presented in Appendix

A.2. A full derivation of the mathematical technique of a separation of variables is beyond the scope of this thesis. A summary is presented in Appendix A.3. The process leads to wave numbers that identify the allowable standing waves in the each of the separated axes, such that a universal modal frequency equation can be defined to predict the frequencies at which these resonances appear.

Standing Waves in a 3D Space

If the acoustic system in considered in 3D then three types of mode exist. *Axial* modes are any path between two parallel rigid reflecting surfaces along the length. *Tangential* modes form at a path between any four reflecting surfaces. *Oblique* modes form at a path between six reflecting surfaces. A combination of all three types of mode will contribute to the the acoustics of a room [41]. Figure 2.4 illustrates the structure of a simple rectangular room of length L in the x-axis, width W in the y-axis and height H in the z-axis.

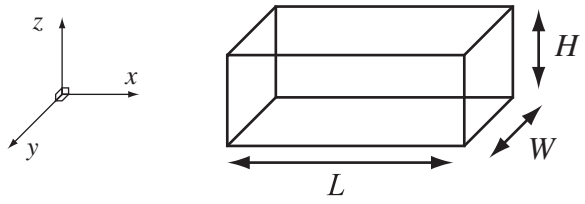


Figure 2.4: *A simple 3D acoustic space in Cartesian coordinates*

Vibrations within the acoustical space will result in resonance. Axial modes exist between each pair of rigid reflecting surfaces. A standing wave is formed such that the molecule displacement component is zero at each boundary. Figure 2.5 illustrates the first two modes in the x-axis between two rigid reflecting surafces at a distance L apart.
Each standing wave exists as a sinusoid evaluated at multiples N of half a wavelength λ_N - at $\frac{\lambda_1}{2} = L$, $\lambda_2 = L$, $\frac{3\lambda_3}{2} = L$, and so on.

$$\lambda_N = \frac{2L}{N} \tag{2.22}$$

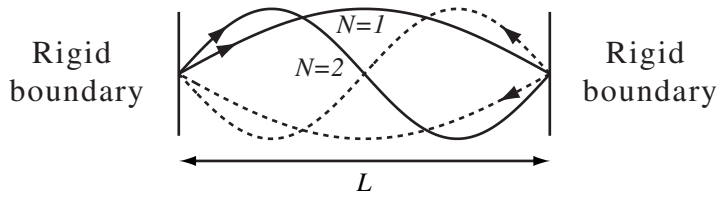


Figure 2.5: *Sinusoidal standing waves at the first two multiples of* $\lambda/2$

Each of these can also be identified with a wave number

$$k = \frac{2\pi}{\lambda_N} \tag{2.23}$$

Substituting for (2.22) into (2.23) gives the wave numbers k_x, k_y and k_z, identifying each of the possible components in the *x*-, *y*- and *z*-axis, respectively, as

$$k_x = \frac{N_x\pi}{L} \qquad N_x = 0, 1, 2, \ldots \tag{2.24}$$

$$k_y = \frac{N_y\pi}{W} \qquad N_y = 0, 1, 2, \ldots \tag{2.25}$$

$$k_z = \frac{N_z\pi}{H} \qquad N_z = 0, 1, 2, \ldots \tag{2.26}$$

The separation of variables method for extrapolating the 3D wave equation into independent terms is demonstrated in Appendix A.3.2. This is achieved by the introduction of the three wave number terms $-k_x^2$, $-k_y^2$ and $-k_z^2$. As in Equation (A.34), the squares of the wave numbers are related in the following manner [43].

$$\frac{\omega^2}{c^2} = k_x^2 + k_y^2 + k_y^2 \tag{2.27}$$

Where ω is the angular frequency of the mode identified by k_x, k_y and k_z. Substituting for (2.24), (2.25) and (2.26) gives

$$\frac{\omega^2}{c^2} = \left(\frac{N_x\pi}{L}\right)^2 + \left(\frac{N_y\pi}{W}\right)^2 + \left(\frac{N_z\pi}{H}\right)^2 \tag{2.28}$$

Rearranging (2.28) and substituting for $\omega = 2\pi f$ leads to the universal modal

frequency equation [45]. This dictates that resonant peaks will appear at values of f_{xyz} in the frequency response of a simple room of length L, width W and height H. The number of the modes are represented by the indices for length $N_x = 0, 1, 2, ...$, width $N_y = 0, 1, 2, ...$ and height $N_z = 0, 1, 2, ...$

$$f_{xyz} = \frac{c}{2}\sqrt{\left(\frac{N_x}{L}\right)^2 + \left(\frac{N_y}{W}\right)^2 + \left(\frac{N_z}{H}\right)^2} \tag{2.29}$$

2.3.5 The Cylindrical Tube

Cylindrical polar coordinates describe a problem-domain in terms of a lengthwise z-axis, a radial r-axis and a rotation about the centre θ, as demonstrated in Figure 2.6. The cylinder has radius a and length L.

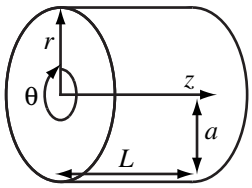


Figure 2.6: *An acoustic tube in cylindrical coordinates*

Using the Laplacian operator for cylindrical coordinates outlined in Appendix A.2, the wave equation becomes

$$\frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} = \frac{1}{r}\frac{\partial p}{\partial r}\left(r\frac{\partial p}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 p}{\partial \theta^2} + \frac{\partial^2 p}{\partial z^2} \tag{2.30}$$

Standing Waves Along a Closed Tube

Note that if pressure variations in the radial r and rotational θ directions is neglected (assumed to be constant), then the planar wave motion along the tube is equivalent to (2.20)

$$\frac{1}{c^2}\frac{\partial^2 p(z,t)}{\partial t^2} = \frac{\partial^2 p(z,t)}{\partial z^2} \tag{2.31}$$

Wave motion in the z-axis is between two parallel rigid reflecting surfaces, and therefore is equivalent to that defined in cartesian space. Standing

waves are formed at multiples of half a wavelength as a result of the positive reflections seen at both closed ends of the tube, as illustrated in Figure 2.7.

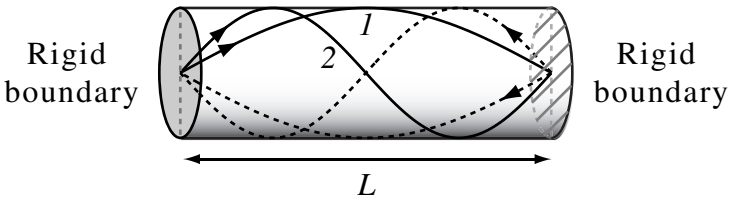


Figure 2.7: *Modes of resonance for $N = 1, 2$ in a cylinder with closed ends*

The first two resonant modes shown highlight the displacement nodes (minima) at both boundaries. Pressure waves, not shown in the diagram, would follow a $\pi/2$ phase lag compared to the displacement components and as such would appear as antinodes (maxima) at either end. These standing waves are sinusoids evaluated at multiples N of π. A wave number for the Nth lengthwise mode can be defined in similar terms to Equation (2.26) as

$$k_z = \frac{N\pi}{L} \qquad N = 1, 2, \ldots \tag{2.32}$$

A separation of variables in cylindrical coordinates is summarised in Appendix A.3.3. This leads to Equation (A.45), describing the relationship between wave numbers k_r and k_z for modes in the r and z axes, respectively, and the standing waves of angular frequency ω, that they represent [43].

$$\frac{\omega^2}{c^2} = k_r^2 + k_z^2 \tag{2.33}$$

Examining only longitudinal wave motion and neglecting terms in r, such that $k_r = 0$, it is possible to equate k_z to the frequencies of the modes of vibration that it represents. Substituting (2.32) into (2.33) gives

$$\frac{\omega^2}{c^2} = \left[\frac{N\pi}{L}\right]^2 \tag{2.34}$$

Substituting for $\omega = 2\pi f$ gives the modal frequencies in the z-axis of a tube of

length L, with two closed ends.

$$f_z = \frac{Nc}{2L} \qquad N = 1, 2, \ldots \tag{2.35}$$

It is worth noting that a similar tube with two open ends would experience negative reflections, resulting in pressure nodes and displacement antinodes at the ends - the opposite of case for the closed tube. The modal frequencies, however, would appear at the same frequencies as specified by (2.35) because the $N\lambda/2$ length of the standing waves remains, despite the reversal of pressure and velocity nodes and antinodes.

Standing Waves Along a Tube Open at One End

A different case where the tube has one closed and one open end is illustrated in Figure 2.8. The velocity components of the first two resonant modes are shown.

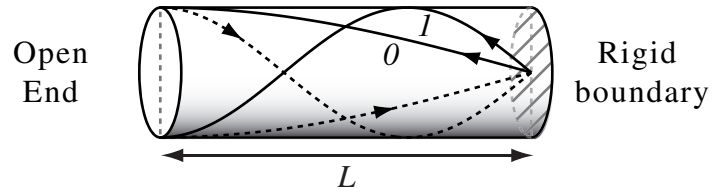


Figure 2.8: *Modes of resonance for $N = 0, 1$ in a cylinder with one open and one closed end*

Reflections exist such that a phase inversion is present at the open end and a phase preservation is seen at the closed end. This gives rise to standing waves with displacement nodes at the closed end, and displacement antinodes at the open end. Modes that satisfy this are at multiples of $2N + 1$ of quarter wavelengths. In other words, at frequencies of wavelength λ_N where $\frac{\lambda_0}{4} = L$, $\frac{3\lambda_1}{4} = L$, $\frac{5\lambda_2}{4} = L$, and so on. The wave number for the Nth lengthwise standing wave is

$$k_z = \frac{(2N+1)\pi}{2L} \qquad N = 0, 1, 2, \ldots \tag{2.36}$$

Examining longitudinal wave motion only, and neglecting r, such that $k_r = 0$, the frequencies of the modes of vibration in the z-axis can be determined.

Substituting (2.36) into (2.33) gives

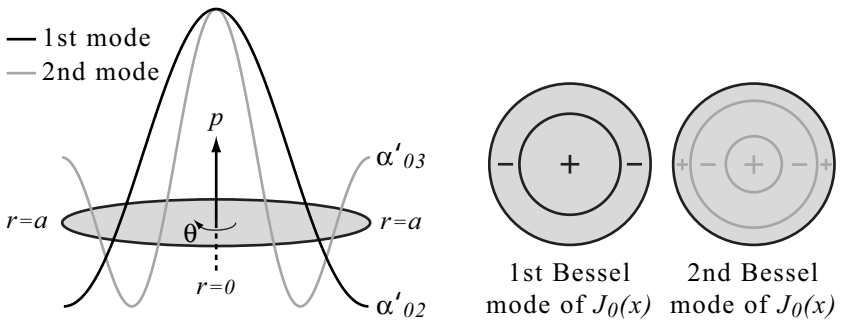
$$\frac{\omega^2}{c^2} = \left[\frac{(2N+1)\pi}{2L}\right]^2 \tag{2.37}$$

Substituting for $\omega = 2\pi f$ gives the modal frequencies along the z axis of a tube with one closed end and one open end.

$$f_z = \frac{(2N+1)c}{4L} \qquad N = 0, 1, 2, \ldots \tag{2.38}$$

Standing Waves Across a Straight Tube

Using a separation of variables, a solution to the cylindrical wave equation in terms of r can be found [46] [47], as demonstrated in Appendix A.3.3. Lengthwise z-axis propagation is separated from the remaining axes and neglected, such that the domain under scrutiny is a circular cross-section of the tube. A Bessel function of the first kind, which is described in Appendix A.3.4, can be used to predict the modes of resonance in the circular plane [48] [43]. As in Equation A.47, pressure standing waves appear in the form of Bessel's function $J_m(k_r r)$ of order m. The function is evaluated out from the centre of the circle at $r = 0$ outwards to the edge at a radius of $r = a$. The wave number k_r identifies each standing wave. Figures 2.9(a) and 2.9(b) show the first two cross-modes of the circle from the side and from above, respectively.



(a) Side view (b) Above view

Figure 2.9: *Pressure modes as a Bessel function of order* $m = 0$

The rotational symmetry about θ that is associated with $m = 0$ can be observed from Figure 2.9(b). As detailed in Appendix A.3.4, points on the Bessel function are annotated such that α'_{mn} is the nth instance of a zero-gradient - in other words, where $J'_m(k_r r) = 0$. The first of such zero-gradients occurs at $r = 0$, where $\alpha'_{01} = k_r 0 = 0$. This corresponds to no transversal modal interactions. The first cross mode (black line) is a Bessel function of order $m = 0$ evaluated from the centre where $r = 0$ and $J_0(k_r r) = 0$, to the first occurrence of $J'_0(k_r r) = 0$ at the tube wall where $r = a$. This happens at $\alpha'_{02} = k_r a = 3.832$ (taken from the graph in Figure A.2). The second cross mode (grey line) is a Bessel function of order $m = 0$, evaluated from the centre to the second occurrence of $J'_0(k_r r) = 0$ at $r = a$. From Figure A.2 $\alpha'_{03} = k_r a = 7.016$. In this way, the wave number in relation to standing waves along the r-axis can be said to be

$$k_r = \frac{\alpha'_{mn}}{a} \qquad m = 0, 1, 2, \ldots \quad n = 0, 1, 2, \ldots \tag{2.39}$$

The frequency of each mode can be obtained with substitution for (2.39) into (2.33) (remembering that neglecting z terms implies that $k_z = 0$).

$$\frac{\omega^2}{c^2} = \left[\frac{\alpha'_{mn}}{a}\right]^2 \tag{2.40}$$

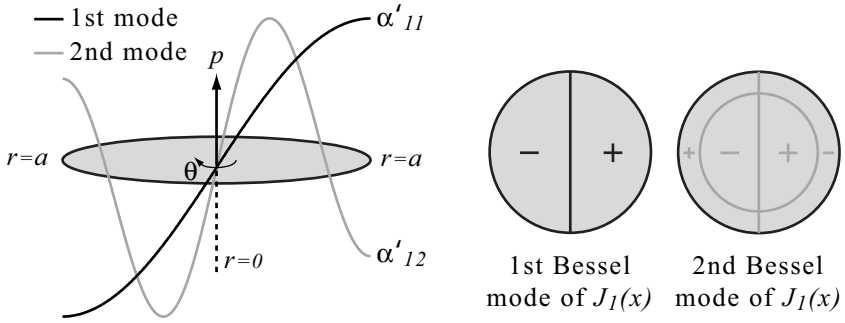
Subsituting for $\omega = 2\pi f$ gives [43]

$$f_{mn} = \frac{c\alpha'_{mn}}{2\pi a} \qquad m = 0, 1, 2, \ldots \quad n = 0, 1, 2, \ldots \tag{2.41}$$

For example, a clarinet has an approximate radius of $a = 8$ mm. With $\alpha'_{02} = 3.832$, the lowest mode of resonance across it with rotational symmetry is $f_{02} = 26.15$ kHz [46]. With $\alpha'_{03} = 7.016$, this second lowest cross-mode is $f_{03} = 47.9$ kHz.

A Bessel function $J_1(k_r r)$ of order $m = 1$ describes a different set of modal resonances across the tube without rotational symmetry. Figure 2.10(a) and 2.10(b) illustrate the first two pressure standing waves across the tube from a side view, and from above, respectively. Note the inverted waveform on

either sides of the centre at $r = 0$.



(a) Side view (b) Above view

Figure 2.10: *Pressure modes as a Bessel function of order $m = 1$*

It is possible to determine the lowest possible cross-mode frequencies without the assumption of rotational symmetry. The first mode (black line) has a zero-gradient at $r = a$, where $\alpha'_{11} = 1.841$ (from Figure A.2). Taking $a = 8$ mm, Equation (2.41) gives the lowest overall cross mode of the clarinet as $f_{11} = 12.6$ kHz. The second (grey line) has a zero-gradient at $\alpha'_{12} = 5.331$ (from Figure A.2). This gives $f_{12} = 36.4$ kHz. Using the same method, a Bessel function with order $m = 2$ has a zero-gradient at $\alpha'_{21} = 3.054$, giving the second lowest mode at $f_{21} = 20.8$ kHz.

These results serve to indicate that, with $f_{11} = 12.6$ kHz, the lowest transverse modal interactions in cylindrical acoustical bores do fall within the range of human hearing.

2.3.6 Spherical Waves in a Conical Tube

A pressure wave in a conical tube, such as that illustrated in Figure 2.11, can be viewed as part of a spherical wave radiating out from an isotropic source at the cone apex [43]. Spherical wave motion follows the *inverse square law*, where the energy present in the wave reduces as inversely proportional to the distance squared from the source r^2. The spherical wave equation also considers angular rotation θ in the *polar* axis, and angular elevation ϕ in the *azimuthal* axis. Using the Laplacian operator for spherical coordinates given

in Appendix A.2, the wave equation is

$$\frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial p}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial \theta}\left(\sin(\theta)\frac{\partial p}{\partial \theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2 p}{\partial \phi^2} \tag{2.42}$$

A spherical wavefront passing through a cone will experience a decay in amplitude with the increase in surface area as it travels along the tube a distance r away from the cone apex. The reduction in amplitude will follow a $1/r$ relationship, given the inverse square law, and that energy is proportional to amplitude squared. The planar wave motion used to define the cylindrical tube assumes no propagational decay and therefore cannot be applied in this case. Spherically symmetrical wave motion in 3D is used to describe the wave p such that it acts only as a function of r, and time t [10]. No considerations are given to the variations with respect to angular rotation θ or ϕ about the source. In other words the wavefront maintains its spherical form as it travels away from the apex and no wave motion takes place across the cone. Therefore, with reference to (2.42), the wave equation for longitudinal motion in a conical tube is

$$\frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial p}{\partial r}\right) \tag{2.43}$$

Standing Waves Along a Conical Tube

Longitudinal standing waves in a conical tube take a similar form as those seen in the closed tube in Figure 2.8. Pressure nodes are formed at the open end and antinodes appear at the apex of the cone. The modes of resonance appear at the same frequencies as described in (2.38). However, the standing wave itself experiences a reduction in amplitude that is inversely proportional to the distance from the cone apex.

2.3.7 The Non-Uniform 1D Tube

The shape formed by an acoustical tube with arbitrarily varying cross-sectional area is quantified in an area function $A(x)$. Wave motion in such a system is not strictly one-dimensional, as the spreading out of a longitudinal

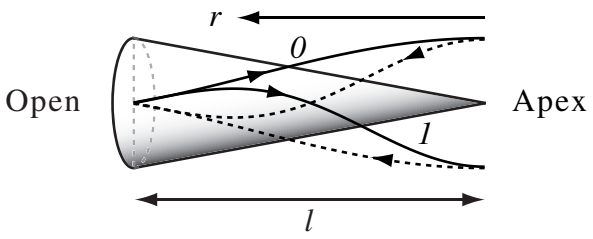


Figure 2.11: *Modes of resonance for* $n = 0, 1$ *in a conical cube*

wavefront as it passes through a sharp increase in tube area would generate a transversal component. However, for a relatively narrow tube that varies smoothly in area along its length, this cross component can be considered negligible. In this case Webster's horn equation [43] is used to describe the longitudinal wave motion

$$\frac{1}{c^2}\frac{\partial^2 p(x,t)}{\partial t^2} = \frac{1}{A(x)}\frac{\partial}{\partial x}\left(A(x)\frac{\partial p(x,t)}{\partial x}\right) \tag{2.44}$$

No assumptions are made regarding the shape of the wavefront and so this equation governs both plane and spherical wave motion derived in Sections 2.3.2 and 2.3.6, respectively.

2.4 Numerical Simulation of the 1D Wave Equation

A simulation of the acoustical properties of a real world system can be accomplished in two stages. Firstly, the identification of an appropriate wave equation that governs the system behaviour allows for a continuous description to be outlined. Assumptions and approximations are made to parts of the model in order to simplify calculations to within the desired scope of the simulation.

Secondly, the system is discretised with respect to the independent variables, time and space. A generalised solution is found which satisfies the system equation at the sample instants and locations. Numerical simulation of the system takes place when calculations that are derived from the solution are performed at the discrete points within the model. Many different

techniques exist with which to discretise, solve and model the equations of a real world system. An exploration of the main types and a thorough treatment of their relation to one another from an acoustical viewpoint can be found in Stefan Bilbao's thesis [9].

In general, time-domain acoustical physical models fall into one of the two following categories: *lumped element* models typically consist of masses, springs and dampers and are used in simulation of vibrations described with an ODE; *distributed systems* model wave propagation in the form of a PDE with the use of transmission lines, or *waveguides*.

2.4.1 Mass and Spring System

The mass and spring system ODE (2.15) has a linear solutions of the form

$$x = e^{-\alpha t} A \cos(\omega_d t + \phi) \tag{2.45}$$

This describes a decaying sinusoid where A is the initial or undamped amplitude of oscillations and the phase shift ϕ depends on the definition of the time $t = 0$.

This type of solution is commonly used as a physical model for a vibrating system. For example, a coupled mass-spring resonator can describe the motion of a clarinet reed [49] or a brass players lips [50] [51]. A network of many interconnected masses and springs can also be used to form a model of a resonating 1D string [52], 2D plate [30] or N-dimensional virtual instrument [53].

2.4.2 Wave Scattering Solution

A numerical solution to the PDE for 1D wave motion exists in the form of the separation of wave variables. No approximation or assumptions are required and so the solution is exact. If the wave equation (2.20) is considered as a difference of two squares then both sides may be factorised

$$\left(\frac{\partial}{\partial x} + \frac{1}{c}\frac{\partial}{\partial t}\right)\left(\frac{\partial}{\partial x} - \frac{1}{c}\frac{\partial}{\partial t}\right) p(x,t) = 0 \tag{2.46}$$

From this it is clear that solutions of the form $p(x \pm ct)$ will satisfy the superposition of waves. This is the D'Alembert exact solution to the 1D wave equation. A thorough treatment of the derivation is presented in Appendix A.1. Using right going p_r and left going p_l components a general form is defined as the sum of travelling wave components

$$p(x,t) = p_r(x-ct) + p_l(x+ct) \tag{2.47}$$

This is the basis for the *waveguide* physical modelling paradigm, which will be covered in greater detail in Section 2.5.

2.4.3 Finite Difference Time-Domain Approximation

A finite difference time-domain (FDTD) scheme may be applied to a wave equation in order to find a solution that is appropriate for numerical simulation [10]. An approximation is made such that each of the differential operators in the wave equation are replaced with a finite difference. A temporal difference can be applied to an ODE. However, where the wave motion is a function of space and time (PDE), additional spatial differences may also be used. In this sense it can be used to construct either a lumped or distributed discrete approximation for the wave equation. For example, a first-order difference in time T for a 1D pressure wave $p(x,t)$ in a distributed model can be defined as

$$\frac{\partial p(x,t)}{\partial t} \approx \frac{p(x,t) - p(x,t-T)}{T} \tag{2.48}$$

Similarly, the difference constructed over a distance X is

$$\frac{\partial p(x,t)}{\partial x} \approx \frac{p(x,t) - p(x-X,t)}{X} \tag{2.49}$$

Second order centred differences follow

$$\frac{\partial^2 p(x,t)}{\partial t^2} \approx \frac{p(x,t+T)-2p(x,t)+p(x,t-T)}{T^2} \tag{2.50}$$

$$\frac{\partial^2 p(x,t)}{\partial x^2} \approx \frac{p(x+X,t)-2p(x,t)+p(x-X,t)}{X^2} \tag{2.51}$$

These may be substituted into the 1D wave equation (2.20)

$$\frac{p(x+X,t)-2p(x,t)+p(x-X,t)}{X^2} = \frac{1}{c^2}\frac{p(x,t+T)-2p(x,t)+p(x,t-T)}{T^2} \tag{2.52}$$

If it is specified that $X = cT$ and the discretisation takes the form $t = nT$ and $x = mX$ then

$$p(m+1,n)-2p(m,n)+p(m-1,n) = p(m,n+1)-2p(m,n)+p(m,n-1) \tag{2.53}$$

For clarity, a time shift of $n-1$ is applied

$$p(m,n) = p(m+1,n-1)+p(m-1,n-1)-p(m,n-2) \tag{2.54}$$

The wave at any point can therefore be approximated by the sum of neighbouring values at one time instant before, minus its own value two time instants before. This scheme is sometimes referred to as a Kirchoff representation [54] because of the summation of actual physical variables, as opposed to the hypothetical travelling components used in the wave scattering solution.

2.4.4 Wave Scattering and FDTD Equivalence

The two spatial discretisation methods that have been presented, wave-scattering and FDTD, can be shown to be equivalent [9] [54]. The FDTD method is seen to satisfy the travelling wave solution if the each of the terms on the right hand side of (2.54) is substituted for the component

decomposition $p(m,n) = p_r(n-m) + p_l(n+m)$ [10].

$$\begin{aligned} p(m,n) &= p_r\left((n-1)-(m+1)\right) + p_l\left((n-1)+(m+1)\right) \\ &+ p_r\left((n-1)-(m-1)\right) + p_l\left((n-1)+(m-1)\right) \\ &- p_r\left((n-2)-m\right) - p_l\left((n-2)+m\right) \qquad (2.55) \\ \implies p(m,n) &= p_r\left(n-m\right) + p_l\left(n-2+m\right) \\ &+ p_r\left(n-2-m\right) + p_l\left(n+m\right) \\ &- p_r\left(n-2-m\right) - p_l\left(n-2+m\right) \qquad (2.56) \end{aligned}$$

This reduces to

$$p(m,n) = p_r\left(n-m\right) + p_l\left(n+m\right) \qquad (2.57)$$

This result is the original left and right going component substitution, therefore the FDTD scheme satisfies the traveling wave solution.

2.4.5 Conical Wave Equation

The equation for longitudinal spherical waves in a conical tube (2.43) can be rewritten as [55]

$$\frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} = \frac{2}{r}\frac{\partial p}{\partial r} + \frac{\partial^2 p}{\partial r^2} \qquad (2.58)$$

This can be rearranged to give the equation for wave motion in a conical tube using spherically symmetrical coordinates [56]

$$\frac{1}{c^2}\frac{\partial^2 rp(r,t)}{\partial t^2} = \frac{\partial^2 rp(r,t)}{\partial r^2} \qquad (2.59)$$

This can be viewed as equivalent to the planar wave PDE (2.20) if the substitution $\psi = rp$ is made

$$\frac{1}{c^2}\frac{\partial^2 \psi}{\partial t^2} = \frac{\partial^2 \psi}{\partial r^2} \qquad (2.60)$$

The d'Alembert solution to the PDE gives the superposition of two bidirectional arbitrary waveforms f_r and f_l

$$\psi(r,t) = f_r(x-ct) + f_l(x+ct) \tag{2.61}$$

Therefore, with substitution for ψ, the wave equation for pressure variations along the conical tube section can be presented in terms of a travelling wave solution

$$p(r,t) = \frac{1}{r}\left[f_r(x-ct) + f_l(x+ct)\right] \tag{2.62}$$

2.4.6 The Webster-Horn Equation

Equation (2.44) describes wave motion in a tube of arbitrary cross-sectional area $A(x)$. However, the dependency on area function means that this wave equation cannot be solved analytically using the separation of travelling wave variables for an exact solution [9]. The smoothly varying tube can be spatially sampled such that it is modelled as a series of short cylindrical or conical sections. Travelling wave solutions are then applicable within each discrete tube section. This notion forms the basis for the piecewise, or concatenated acoustic tube model [18]. A digital waveguide is used to represent the wave motion within each short tube section, through a discrete solution or approximation to the wave equation. A connected chain of waveguides acts to approximate the overall behaviour of the continuous tube.

2.5 The 1D Digital Waveguide

The main body of the work in this thesis is focused on the use of the digital waveguide as the fundamental component in a distributed acoustical model. It uses the wave scattering solution to the 1D PDE presented in Section 2.4.2 to simulate wave propagation. The development of much of the waveguide theory is accredited to work done by Julius Orion Smith III at CCRMA, Stanford University, whose online book [10] is an extensive source

for reference.

2.5.1 Bi-Directional Wave Decomposition

Simulation of acoustic wave propagation in a transmitting medium is founded on the d'Alembert solution (2.47) to the 1D wave equation. Waveguide modelling theory uses the notion that 1D wave motion may be considered a composition of left going and right going components. The instantaneous magnitude of the oscillations can then be obtained as a sum of left and right components at any point along the modelled medium. It is worth noting that the magnitude of vibrations is the physical variable under simulation that would be observed in the real world system. The bidirectional travelling wave components are a hypothetical consideration to facilitate propagation.

For example, an acoustic tube can be discretised in both time and space to give a number of sample reference points, each separated by a *bi-directional digital delay*. Figure 2.12 illustrates how this applies to a model of pressure variations in an unbounded 1D homogenous medium, such as an infinitely long tube of constant cross-sectional area (and hence constant impedance). Sampling instants, indexed with n are separated by the delay units, marked z^{-1}.

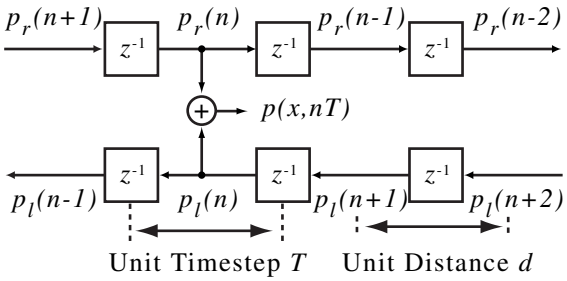


Figure 2.12: *Pressure components in a 1D waveguide acoustic tube model*

The dissection of the medium into digital waveguides requires a discrete version of the d'Alembert solution. The magnitude p of the pressure variations at a time nT and at a distance x along the tube is the sum of the

right p_r, and left p_l going components [33]

$$p(x, nT) = p_r(x - cnT) + p_l(x + cnT) \tag{2.63}$$

Wave simulation in (2.63) describes exact lossless non-dispersive 1D signal propagation. The simplicity and accuracy of the propagational methodology make it highly suitable for computer numerical simulation.

The delay units transfer signal values from one spatial sampling point to another in both the right and left directions. Each therefore represents both a length d of each discrete piece of tube and also a unit time-step T. These physical quantities within the model are directly related to the speed of sound c, such that the sampling frequency f_s is

$$f_s = \frac{1}{T} = \frac{c}{d} \tag{2.64}$$

2.5.2 Scattering at an Admittance Discontinuity

The physical modelling representation of an acoustic tube of constantly varying cross-sectional area is of particular interest to this work. The area function is spatially sampled such that it is modelled as a series of adjoining cylindrical tube elements. The *Kelly-Lochbaum* (KL) signal processing unit is used to simulate wave scattering at the junction between two tubes of different cross-sectional area, and therefore of different admittance. The KL junction will now be examined following the notation of some of the detailed derivations in the literature [18], [56], [46], [57].

Figure 2.13 illustrates the various signals surrounding an impedance discontinuity between tube sections i and $i+1$. The diagram also demonstrates the introduction of the KL scattering junction in between the bidirectional unit delay elements seen in Figure 2.12. It is convenient at this stage to annotate pressure components as inputs and outputs to and from the junction rather than according to their direction of travel. Therefore pressure p_i^+ and velocity u_i^+ represent inputs to the KL junction from tube section i. Similarly, pressure p_i^- and velocity u_i^- correspond to outputs from the junction towards

i. All signal components are between the two unit delays and therefore exist at the same time index *n*.

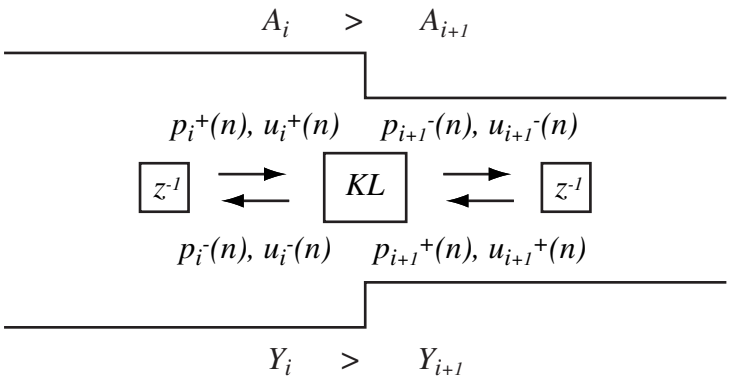


Figure 2.13: *Signal scattering at an admittance discontinuity*

Continuity laws dictate that the pressure and velocity remain constant about the junction between i and $i+1$. Therefore:

- The total pressure at waveguide i is always the sum of the input and output components as defined in (2.63):

$$p_i = p_i^+ + p_i^- \tag{2.65}$$

- The instantaneous pressure at each connection is equal, and can therefore be referred to using the singular junction pressure term p_J. The input and output pressure components will also follow this relationship:

$$p_i = p_{i+1} = p_J \quad \Longrightarrow \quad p_i^+ + p_i^- = p_{i+1}^+ + p_{i+1}^- \tag{2.66}$$

- The total velocity is the sum of the two components, with outputs notated to be inverted with respect to inputs due to the reversed direction of flow:

$$u_i = u_i^+ - u_i^- \tag{2.67}$$

- The instantaneous net flow is zero. Therefore the input and output

velocities also remain balanced:

$$u_i + u_{i+1} = 0 \quad \Longrightarrow \quad u_i^+ - u_i^- + u_{i+1}^+ - u_{i+1}^- = 0 \tag{2.68}$$

The KL junction acts to scatter approaching signals according to the change in impedance. As in (2.7), the characteristic impedance of tube section i is the ratio of its pressure and velocity components. Hence, the input and output velocity components to and from i in terms of acoustic admittance and pressure are

$$u_i^+ = Y_i p_i^+ \qquad\qquad u_i^- = Y_i p_i^- \tag{2.69}$$

Considering (2.67), the total velocity on connection i is

$$u_i = u_i^+ - u_i^- = Y_i \left(p_i^+ - p_i^- \right) \tag{2.70}$$

Substituting (2.70) into (2.68) gives the total velocity about a junction in pressure terms only.

$$Y_i \left(p_i^+ - p_i^- \right) + Y_{i+1} \left(p_{i+1}^+ - p_{i+1}^- \right) = 0 \tag{2.71}$$

Given (2.65), substitution of $p^- = p_J - p^+$ into (2.71) leads to

$$Y_i \left(2p_i^+ - p_J \right) + Y_{i+1} \left(2p_{i+1}^+ - p_J \right) = 0 \tag{2.72}$$

Rearranging this for p_J gives the *two-port scattering equation* for junction pressure in terms of the sum of its input pressure components, scaled by the step in admittance between i and $i+1$.

$$p_J = \frac{2 \left(Y_i p_i^+ + Y_{i+1} p_{i+1}^+ \right)}{Y_i + Y_{i+1}} \tag{2.73}$$

Calculation of junction pressure is an intermediate step in determining junction outputs in the scattering algorithms. Pressure value p_J itself is only directly of interest where the junction is under scrutiny, either as an audio output from the system, or for graphical drawing purposes to illustrate wave

motion. In many cases it is more economical in computational terms to reduce the scattering equations to pressure output terms only, with as few multiplications as possible. Scattering equations which bypass consideration of p_J are required. Outputs from the junction p_i^- and p_{i+1}^- can then be derived substituting for p_J in (2.65)

$$\begin{aligned} p_i^- &= p_J - p_i^+ = \frac{Y_i - Y_{i+1}}{Y_i + Y_{i+1}} p_i^+ + \frac{2Y_{i+1}}{Y_i + Y_{i+1}} p_{i+1}^+ \\ p_{i+1}^- &= p_J - p_{i+1}^+ = \frac{2Y_i}{Y_i + Y_{i+1}} p_i^+ - \frac{Y_i - Y_{i+1}}{Y_i + Y_{i+1}} p_{i+1}^+ \end{aligned} \tag{2.74}$$

A reflection coefficient between the two impedances can be formed using either admittance Y, impedance Z or, given (2.11), tube cross-sectional area A

$$r = \frac{Y_i - Y_{i+1}}{Y_i + Y_{i+1}} = \frac{Z_{i+1} - Z_i}{Z_i + Z_{i+1}} = \frac{A_i - A_{i+1}}{A_i + A_{i+1}} \tag{2.75}$$

This leads to simplified scattering equations for the outputs from a KL junction in input terms only [56]

$$\begin{aligned} p_i^- &= r p_i^+ + (1-r) p_{i+1}^+ \\ p_{i+1}^- &= (1+r) p_i^+ - r p_{i+1}^+ \end{aligned} \tag{2.76}$$

Therefore some amount of signal incident upon the KL junction from either direction is transmitted through, and some is reflected back. Finally, for speed and simplicity of computation, it is possible to derive forms of (2.76) using only one multiplication. Further substitution is made for the intermediate signal $w = r\left[p_i^+ - p_{i+1}^+\right]$.

$$\begin{aligned} p_i^- &= p_{i+1}^+ + w \\ p_{i+1}^- &= p_i^+ + w \end{aligned} \tag{2.77}$$

Schematic representations for the signal flow in the KL scattering junction for (2.76) and (2.77) are shown in Figures 2.14(a) and 2.14(b), respectively.

The remainder of this work will focus solely on waveguide theory using pressure signals, however, the above derivation for KL signal scattering at

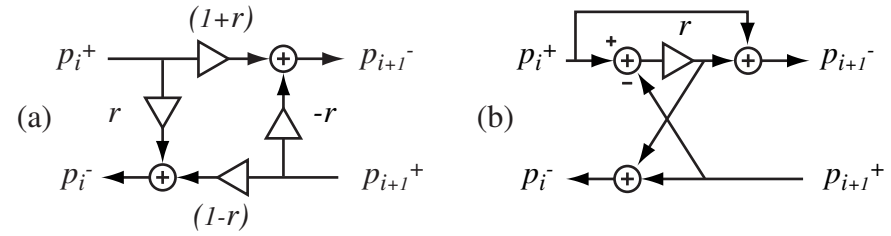


Figure 2.14: *KL scattering of pressure signals in (a) the two-port junction and (b) the one-multiply equivalent*

an impedance discontinuity can be derived for volume velocity in a similar manner. Elimination of pressure signals from (2.66), and the definition of a reflection coefficient, results in the following velocity outputs from the junction [56].

$$\begin{aligned} u_i^- &= ru_i^+ + (1+r)u_{i+1}^+ \\ u_{i+1}^- &= (1-r)u_i^+ - ru_{i+1}^+ \end{aligned} \tag{2.78}$$

Similarly, the one-multiply equivalents with the intermediate signal $w = r\left[u_i^+ + u_{i+1}^+\right]$ are

$$\begin{aligned} u_i^- &= u_{i+1}^+ + w \\ u_{i+1}^- &= u_i^+ - w \end{aligned} \tag{2.79}$$

Figure 2.15 details the signal flow schematic for acoustic volume velocity about a KL scattering junction.

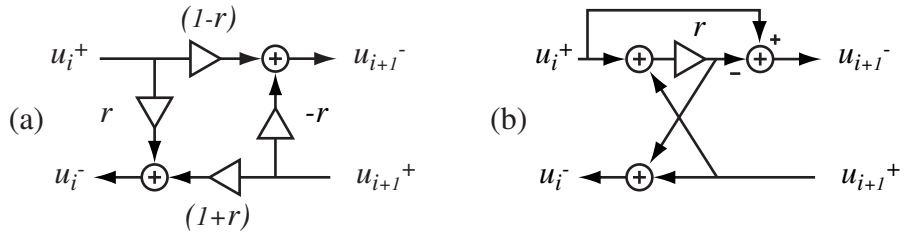


Figure 2.15: *KL scattering of volume velocity signals in (a) the two-port junction and (b) the one-multiply equivalent*

2.5.3 Simple Reflection Boundary Implementation

The 1D boundary junction equation is derived by considering the change in admittance experienced across the connections of a simple terminating junction [58] as illustrated in Figure 2.16. Additional notation has been included in the diagram to indicate which of the neighbouring junctions a pressure signal relates to. For example air pressure values labelled $p^+_{B,J}$ indicate an incoming pressure at junction B from junction J (at J a unit time step before), and those labelled $p^-_{B,J}$ show the outgoing pressure at B, to J (reaching J a time step later).

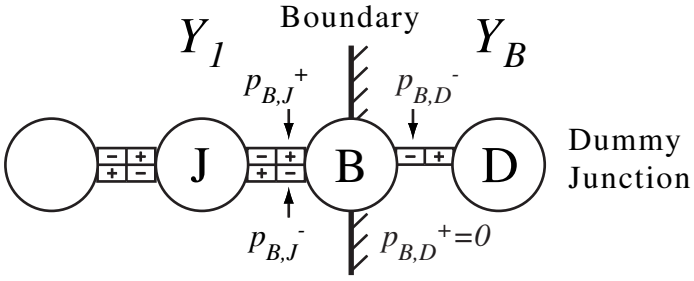


Figure 2.16: *The general one-connection boundary junction*

A dummy junction D is inserted within the bounding medium beyond the actual boundary junction B such that it is not apparent to the main body of the waveguide chain. Such a configuration is termed the one-connection boundary junction as the second output from B to D within the bounding surface is a hypothetical consideration used in the derivation. Junction D does not contribute energy back towards the waveguide chain, acting only to absorb a proportion of the pressure incident upon junction J, therefore $p^+_{B,D} = 0$. A reflection coefficient r is defined such that a proportion of the energy incident upon B from J is reflected back towards J.

$$p^-_{B,J} = r p^+_{B,J} \tag{2.80}$$

If this boundary represents a raditaing surface, such as the open end of a pipe,

then the output from the system is the remainder of of the reflected signal

$$p^{-}_{B,D} = (1-r)\, p^{+}_{B,J} \tag{2.81}$$

As described in (2.75), an impedance discontinuity between two mediums, Y_1 and Y_B constitutes a reflection coefficient $r = \frac{Y_1 - Y_B}{Y_1 + Y_B}$. The two-port scattering equation takes the form

$$p_B = \frac{2\left(Y_1 p^{+}_{B,J} + Y_B p^{+}_{B,D}\right)}{Y_1 + Y_B} \tag{2.82}$$

We define the ratio between the two admittances as $Y_B = \mu Y_1$, where $\mu = \frac{1-r}{1+r}$, and eliminate $p^{+}_{B,D} = 0$. The pressure on the boundary junction is

$$p_B = \frac{2Y_1 p^{+}_{B,J}}{Y_1 + \mu Y_1} = \frac{2p^{+}_{B,J}}{1+\mu} \tag{2.83}$$

Substituting for r

$$p_B = \frac{2p^{+}_{B,J}}{1 + \frac{1-r}{1+r}} \tag{2.84}$$

Finally, rearranging for the general form of the scattering equation for the pressure at a one-connection boundary node, gives

$$p_B = (1+r) p^{+}_{B,J} \tag{2.85}$$

2.5.4 The Conical Waveguide

The concatenated acoustic tube model can also be formulated using conical waveguide segments [59]. Figure 2.17 illustrates the junction formed between two conical tube elements, *a* and *b*.

The junction is defined such that the two cones intersect at a point where their cross-sectional areas are equal, as indicated by the cross-hatched region *A*. Although the two tube segments are conical in form, the apex of each cone appears only for illustrative and derivational purposes and is not actually present in the model. Each cone therefore comprises a unit conical waveguide of length d and a theoretical tip of length r, measured as the distance from

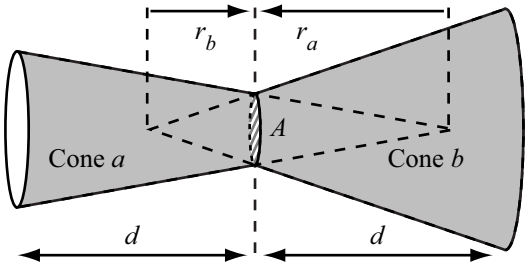


Figure 2.17: *A junction of two conical tube elements*

the effective cone apex to the junction at A. The pressure on each conical waveguide is governed by the travelling wave solution defined in (2.62). If we define a value α for the junction as

$$\alpha = \frac{c(r_a - r_b)}{2r_a r_b} \tag{2.86}$$

It can be shown that a frequency dependent reflection exists at the junction [56] as

$$R(\omega) = -\frac{\alpha}{j\omega + \alpha} \tag{2.87}$$

A discrete form of this reflection function can be identified and constructed as a digital filter. This replaces the reflection constant in the standard KL junction shown in Figure 2.14.

The conical junction model does, however, present some stability issues in certain configurations [56]. Smoothly varying acoustic tube models can be generated with overall stability. However, particular problems emerge in the simulation of a quickly diverging flared horn. The conical waveguide is defined under the assumption that spherical wave motion is maintained throughout the tube section. That is to say that wavefronts at different positions within it are parallel. A wavefront passing through a junction between two cones, however, naturally experiences a change in shape. Once it has entered the new cone, the wavefront will be spherical within the new coordinates system, but not parallel to those in the previous system. This concept is illustrated in Figure 2.18, where a junction between two cone segments, a and b, is shown. They are arranged such that they form a non-convex junction where both diverge in the same direction, with cone a at a

lower rate than b.

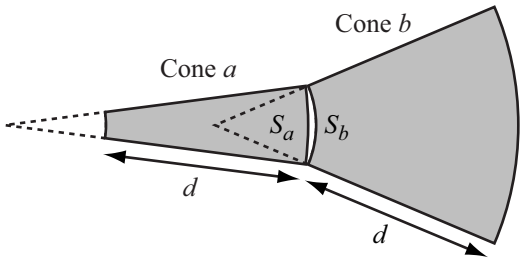


Figure 2.18: *Missing volume in a non-convex conical tube junction*

Notations S_a and S_b indicate a wavefront just before leaving cone a and just after entering cone b, respectively. The two are clearly not parallel. The step made betweens the two cone sections, apparent as the white region, has been identified as a missing volume [60]. Instabilities in the conical junction come about because wave motion within this region is not accounted for. A reverse situation also exists for a convex junction, where a doubly defined volume is problematic. This problem may be addressed with the introduction of hyperbolic waveguides. The method uses a change of coordinates to convert Webster's equation (2.44) into a Schrödinger form for discretisation [60]. It is of particular interest in the modelling of the flared bell at the end of a brass instrument [61].

2.5.5 The Fractional Delay

The discretisation that is chosen for solving or approximating the wave equation in a system defines the spatial sampling instants. Valid output from the system can be obtained from one of these finite points at the waveguide junctions, where a physical variable exists. It is possible to determine the signal value in between the sampling instants with the use of a fractional delay filter [56]. This allows for fine tuning of the sampling instants such that bandlimited interpolation can be used to evaluate a signal at an arbitrary point in time or space. There is, however, an increase in computational requirements of a system employing such additional filter units.

2.5.6 Completing the 1D Physical Model

Once the propagational mechanism is defined, additional factors can be introduced to increase the depth of representation in the model. Characteristics of the medium, such as losses and dispersion that were originally omitted from the derivation of the wave equation can be reintroduced. The linear nature of the waveguide allows for the effect of such properties to be commuted from each discrete unit into a singular digital filter at either end. This process separates the ideal propagational model from reflections and nonlinearities and results in a more efficient computational model. A frequency dependant reflection, such as that observed at the opening of a many wind instruments, may be also be accounted for in the bounding digital filter. For example, a simple physical model of a clarinet consists of three distinct parts; propagational, reflectance and nonlinear excitation [36]. This is illustrated in Figure 2.19.

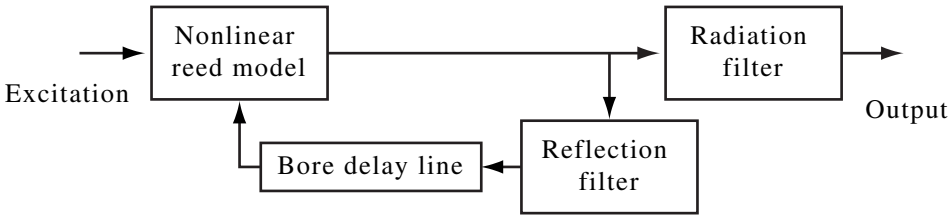


Figure 2.19: *A physical model of a clarinet*

The central acoustic bore is represented in 1D idealistic propagational terms as a simple delay line. In this case, the bidirectional delays within the waveguides perform only time-step operations and so have been collected together into one equivalent unidirectional delay line. Reflectance and radiation filters model the waveform interaction with the clarinet bell. The excitation to the clarinet is generated with a constant flow of pressure input to a simple mass and spring resonator.

2.6 The Digital Waveguide Mesh

The waveguide theory discussed in Section 2.5 can be extended for use in the modelling of wave propagation in structures with increased dimensional representation. This requires a lattice of multiple-port scattering junctions, each formed where a number of waveguides meet at the same spatial sampling point. In an arbitrary configuration, this would constitute a Digital Waveguide Network (DWN) [10]. More specifically, where junctions are placed at regular intervals in a grid system, the model is called a Digital Waveguide Mesh (DWM). The arrangement of junctions and the number of connections at each defines how the model represents the target structure. For example, a three-port scattering junction can be used to simulate signal interaction at a point where two 1D systems intersect. This could be a side branch in an acoustic tube such as a tone-hole in a musical instrument, or the connection of the nasal cavity to the main tract in a vocal model. A three-port (or higher) junction might also be used to model a point where three (or more) waveguides meet on a 2D DWM model of vibrating plane, such as a drum skin [12]. Similarly, an N-port junction might be employed where many waveguides meet in a 3D DWM model of, for example, the acoustics of a room [13].

2.6.1 General Multiple-Port Scattering

The multiple port scattering junction describes the summing and distribution of N incoming and outgoing pressure waves incident at the same temporal and spatial sampling location. Figure 2.20(a) illustrates the unit-junction with N connections and Figure 2.20(b) extends this to include surrounding junctions.

Lossless wave propagation through the junction is maintained by ensuring that the multi-port equivalences for the continuity laws outlined in Section 2.5.2 are adhered to for the N-connections:

- The pressure at the ith waveguide connection to junction J is the sum of

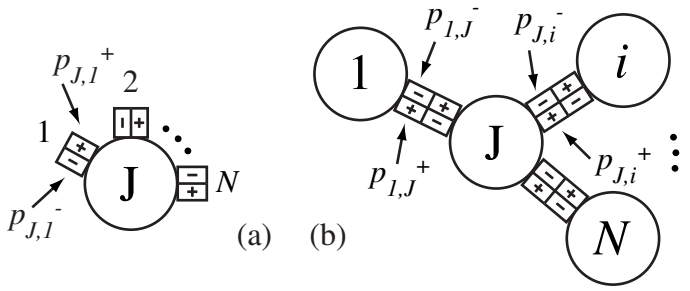


Figure 2.20: *N-port scattering: (a) the unit junction, and (b) with N neighbouring junctions*

the incoming and outgoing pressure components to that connection:

$$p_{J,i} = p_{J,i}^{+} + p_{J,i}^{-} \tag{2.88}$$

- The sum of the input velocities is equal to the sum of the output velocities (the net flow is equal to zero):

$$\sum_{i=1}^{N} u_{J,i}^{+} = \sum_{i=1}^{N} u_{J,i}^{-} \tag{2.89}$$

- The pressure at each waveguide connection to a junction is equal and can be referred to using the singular pressure term p_J:

$$p_1 = p_2 = p_i = \cdots = p_N = p_J \tag{2.90}$$

As with the derivation for KL scattering, the relationship between pressure, velocity and the admittance of the propagating medium described in (2.7) can be extended to include input and output considerations. Velocity components on each waveguide i as described in (2.69) can be summed around the N connections in a manner that coincides with continuity law (2.89)

$$\sum_{i=1}^{N} Y_i p_{J,i}^{+} = \sum_{i=1}^{N} Y_i p_{J,i}^{-} \tag{2.91}$$

Rearranging (2.88) to give $p_{J,i}^{-} = p_J - p_{J,i}^{+}$ and substituting into (2.91) to

eliminate output components, gives

$$\sum_{i=1}^{N} Y_i p_{J,i}^{+} - \sum_{i=1}^{N} Y_i p_J + \sum_{i=1}^{N} Y_i p_{J,i}^{+} = 0 \tag{2.92}$$

Hence

$$\sum_{i=1}^{N} Y_i p_J = 2 \sum_{i=1}^{N} Y_i p_{J,i}^{+} \tag{2.93}$$

This leads to the general form of the *N-port scattering equation*. The pressure p at a junction J in terms of input pressures from N connecting waveguides of admittance Y_i is

$$p_J = 2 \frac{\sum_{i=1}^{N} Y_i p_{J,i}^{+}}{\sum_{i=1}^{N} Y_i} \tag{2.94}$$

For two port scattering (2.76) minimal terms were derived for efficient algorithms that bypassed p_J. A similar method may be used to examine signal flow at a multiple port junction. The scattering equations may be condensed into terms concerning inputs and outputs only. This premise is well discussed by way of example before a general case is outlined. Consider the junction in Figure 2.21 where three waveguides of admittance Y_i meet [56].

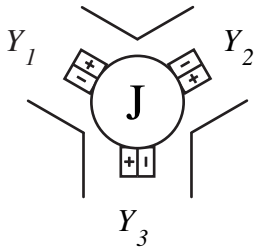


Figure 2.21: *A three-port junction*

The scattering equation can be written, with $N = 3$, as

$$p_J = \frac{2\left(Y_1 p_1^{+} + Y_2 p_2^{+} + Y_3 p_3^{+}\right)}{Y_1 + Y_2 + Y_3} \tag{2.95}$$

Each of the three outputs are equal to the junction pressure minus the input

from the same branch, as in (2.88)

$$\begin{aligned} p_1^- &= p_J - p_1^+ = \frac{[Y_1 - Y_2 - Y_3]\, p_1^+ + 2Y_2 p_2^+ + 2Y_3 p_3^+}{Y_1 + Y_2 + Y_3} \\ p_2^- &= p_J - p_2^+ = \frac{[Y_2 - Y_1 - Y_3]\, p_2^+ + 2Y_1 p_1^+ + 2Y_3 p_3^+}{Y_1 + Y_2 + Y_3} \\ p_3^- &= p_J - p_3^+ = \frac{[Y_3 - Y_1 - Y_2]\, p_3^+ + 2Y_1 p_1^+ + 2Y_2 p_2^+}{Y_1 + Y_2 + Y_3} \end{aligned} \tag{2.96}$$

The reflection coefficient seen at each branch is related to its own admittance, and that of the remaining branches

$$r_1 = \frac{Y_1 - Y_2 - Y_3}{Y_1 + Y_2 + Y_3} \qquad r_2 = \frac{Y_2 - Y_1 - Y_3}{Y_1 + Y_2 + Y_3} \qquad r_3 = \frac{Y_3 - Y_1 - Y_2}{Y_1 + Y_2 + Y_3} \tag{2.97}$$

Substitution for (2.97) into (2.96) gives the output pressure components for the three port junction in reduced mathematical terms. It is clear that each output receives a reflection from the same connection, and some amount of transmitted signal from all of the others. The proportion of incident energy reflected/transmitted at each branch is determined by the relationship between the admittances in (2.97).

$$\begin{aligned} p_1^- &= r_1 p_1^+ + (1 + r_2)\, p_2^+ + (1 + r_3)\, p_3^+ \\ p_2^- &= r_2 p_2^+ + (1 + r_1)\, p_1^+ + (1 + r_3)\, p_3^+ \\ p_3^- &= r_3 p_3^+ + (1 + r_1)\, p_1^+ + (1 + r_2)\, p_2^+ \end{aligned} \tag{2.98}$$

A general form for the output components from a junction with N differing admittance connections can be defined in a similar manner. The general scattering equation in admittance terms is

$$p_J = \frac{2\sum_{i=1}^{N} Y_i p_{J,i}}{\sum_{i=1}^{N} Y_i} \tag{2.99}$$

From (2.88), pressure output onto any connection k is

$$p_k^- = p_J - p_k^+ = \frac{2\sum_{i=1}^{N} Y_i p_{J,i}^+}{\sum_{i=1}^{N} Y_i} - p_k^+ \tag{2.100}$$

This leads to the output pressure component on any connection k in terms of

the input from k, and all from other inputs $0 < i < N$, exclusive of k

$$p_k^- = \frac{\left[2Y_k - \sum_{i=1}^{N} Y_i\right] p_k^+ + 2\sum_{i=1, i\neq k}^{N} Y_i p_{J,i}^+}{\sum_{i=1}^{N} Y_i} \tag{2.101}$$

As with the three port example, this can be viewed as each output being a proportion of input reflected from the same connection, plus some amount of each of the other inputs transmitted through the junction. Splitting (2.101) into reflected and transmitted parts

$$p_k^- = \frac{\left[2Y_k - \sum_{i=1}^{N} Y_i\right]}{\sum_{i=1}^{N} Y_i} p_k^+ + \frac{2\sum_{i=1, i\neq k}^{N} Y_i p_{J,i}^+}{\sum_{i=1}^{N} Y_i} \tag{2.102}$$

As such we can define a reflection coefficient for the *kth* connection as

$$r_k = \frac{2Y_k - \sum_{i=1}^{N} Y_i}{\sum_{i=1}^{N} Y_i} \tag{2.103}$$

Then the output component seen at each connection will be that reflected from the same branch, plus all other signals transmitted through

$$p_k^- = r_k p_k^+ + \sum_{i=1, i\neq k}^{N} (1 + r_i)\, p_{J,i}^+ \tag{2.104}$$

2.6.2 Multiple-Port Scattering with Equal Admittance

In the case where the junction is defined at the connection of N equal impedance waveguides, such as might be seen in a DWM of a homogenous medium, the scattering equations may be further condensed. Scattering at junction J with N equal admittance connecting waveguides $Y_1 = Y_2 = \cdots = Y_N$ reduces (2.94) to

$$p_J = \frac{2}{N} \sum_{i=1}^{N} p_{J,i}^+ \tag{2.105}$$

Once the pressure at each junction has been calculated, the output to each waveguide is set by Equation 2.88 as

$$p_{J,i}^- = p_J - p_{J,i}^+ \tag{2.106}$$

An increment in the time index is implemented by transferral of all outputs at junction J to the relative inputs of the neighbouring junctions. The input to the neighbouring junction i is then the time-delayed output from junction J.

$$p^+_{J,i} = z^{-1} p^-_{i,J} \tag{2.107}$$

It is useful to collect the three important equations (2.105), (2.106) and (2.107) together in this manner as collectively they constitute a single and complete pass of the scattering algorithms as applied to the main body of a DWM model. During such an operation the pressure at each junction is directly calculated from the inputs, followed by each output, ready to be transferred to the neighbouring junction on that connection for the next timestep.

2.6.3 The General Multiple-Port Boundary Junction

The nature of an N-port boundary junction is considered in a similar manner to the one-connection case. Figure 2.22 illustrates the dummy junction D in the bounding medium of admittance Y_B, the actual boundary junction B and the N connections to neighbouring junctions, each of waveguide admittance Y_i.

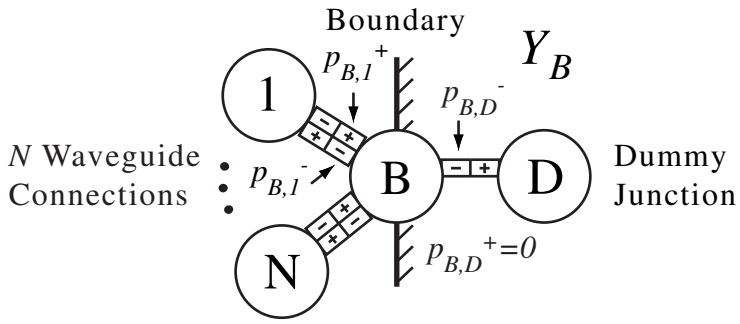


Figure 2.22: *The N-connection waveguide boundary junction*

Taking into account the additional connection to D, and remembering that $p^+_{B,D} = 0$, the general scattering equation takes the following form.

$$p_B = \frac{2\sum_{i=1}^{N} Y_i p^+_{B,i}}{\sum_{i=1}^{N} Y_i + Y_B} \tag{2.108}$$

The output on the *kth* branch can be determined in the same way as (2.100).

$$p_k^- = p_B - p_k^+ = \frac{2\sum_{i=1}^{N} Y_i p_{B,i}^+}{\sum_{i=1}^{N} Y_i + Y_B} - p_k^+ \tag{2.109}$$

This leads to

$$p_k^- = \frac{\left[2Y_k - \sum_{i=1}^{N} Y_i - Y_B\right] p_k^+ + 2\sum_{i=1, i\neq k}^{N} Y_i p_{B,i}^+}{\sum_{i=1}^{N} Y_i + Y_B} \tag{2.110}$$

The reflection coefficient for each connection is defined, taking into account the boundary admittance, as

$$r_k = \frac{2Y_k - \sum_{i=1}^{N} Y_i - Y_B}{\sum_{i=1}^{N} Y_i + Y_B} \tag{2.111}$$

The signal transferred through the connection is no longer simply $(1 + r_k)$ because of the additional Y_B term. If we define the adjusted transmission coefficient τ_k to be

$$\begin{aligned} \tau_k &= 1 + r - \frac{Y_B}{\sum_{i=1}^{N} Y_i + Y_B} \\ &= \frac{2\sum_{i=1, i\neq k}^{N} Y_i}{\sum_{i=1}^{N} Y_i + Y_B} \end{aligned} \tag{2.112}$$

Each output from the junction can then be considered as a reflection from the same port plus some amount transmitted through from each of the remaining connections.

$$p_k^- = r_k p_k^+ + \sum_{i=1, i\neq k}^{N} \tau_i p_{B,i}^+ \tag{2.113}$$

2.6.4 The Multiple-Port Boundary Junction with Equal Admittances

If the N connections in Figure 2.22 represent waveguides in a homogenous medium, such as in a DWM used to model wave propagation in air, then all are of equal admittance $Y_1 = Y_2 = \cdots = Y_N$. All can be replaced with a singular mesh admittance term Y_m. N-port scattering can now be derived in a similar manner to the one-connection case in Figure 2.16. The ratio between

the admittances is $Y_B = \mu Y_m$, where $\mu = \frac{1-r}{1+r}$ such that a reflection coefficient $r = \frac{Y_m - Y_B}{Y_m + Y_B}$ exists between the mesh and the boundary. The junction pressure is calculated from the N-port scattering equation (2.94) in terms of pressure inputs and singular mesh admittance Y_m

$$p_B = \frac{2\sum_{i=1}^{N} Y_m p_{B,i}^{+}}{\sum_{i=1}^{N} Y_m + \mu Y_m} \tag{2.114}$$

Eliminating Y_m, the pressure at the boundary is

$$p_B = \frac{2\sum_{i=1}^{N} p_{B,i}^{+}}{(N+\mu)} \tag{2.115}$$

As $\mu = \frac{1-r}{1+r}$, the general form of the scattering equation for the boundary junction B, at a medium providing a reflection r, with N mesh-body neighbours is

$$p_B = \frac{2\sum_{i=1}^{N} p_{B,i}^{+}}{(N+\frac{1-r}{1+r})} \tag{2.116}$$

This results in the following scattering equations for a boundary junction of N connections, with $N = 1, 2, 3, 4$. Confirmation of this derivation method can be obtained by comparison of the equation for $N = 1$ and the one-connection case (2.85).

- 1-Port Bounding Node: $p_B = (1+r)p_{B,1}^{+}$
- 2-Port Bounding Node: $p_B = \frac{2(1+r)}{3+r}(p_{B,1}^{+} + p_{B,2}^{+})$
- 3-Port Bounding Node: $p_B = \frac{(1+r)}{2+r}(p_{B,1}^{+} + p_{B,2}^{+} + p_{B,3}^{+})$
- 4-Port Bounding Node: $p_B = \frac{2(1+r)}{5+3r}(p_{B,1}^{+} + p_{B,2}^{+} + p_{B,3}^{+} + p_{B,4}^{+})$

2.6.5 Junction Excitation

Input to the waveguide structure is achieved with the introduction of an additional junction connection. This can be established in either a regular scattering or boundary junction. The scattering equation for any junction J with N standard waveguide ports and an external connection contributing

p^+_{ext} with wave admittance Y_{ext}, is

$$p_J = 2\frac{\sum_{i=1}^{N} Y_i p^+_{J,i} + Y_{ext} p_{ext}}{\sum_{i=1}^{N} Y_i + Y_{ext}} \tag{2.117}$$

This excitation can be applied to any number of junctions. Input to one node constitutes a point source, whereas input to a line or plane of multiple nodes represents a source with physical dimensions.

2.6.6 2D Mesh Topology and Dispersion Effects

An important factor in the construction of a DWM is the manner in which the waveguides are arranged to fill the modelled space. As a 1D propagational model the waveguide is an exact representation, in that it provides a complete solution to the governing 1D PDE. A model that uses waveguides connected in a grid form to extend the dimensional representation loses exactitude. Considered in 2D, a wavefront emanating from a source propagates as an expanding circle. True 2D waveform simulation would require that an infinite number of plane waves radiates out from the source. Distributed evenly amongst all angles, the waves would combine as a circular front. A discrete model that uses a finite number of 1D waveguides to simulate 2D propagation will therefore present an approximation to the inverse square law bound circular spreading. Errors resulting from this approximation manifest as a direction and frequency dependent *dispersion*. The extent to which this effect is of concern depends on the arrangement of the junctions and the time step represented in each waveguide. The sampling frequency f_s of the N dimensional mesh that uses waveguides of length d, and supports a wave speed c is [62]

$$f_s = \frac{c\sqrt{N}}{d} \tag{2.118}$$

The Rectilinear Mesh

The originally proposed rectilinear topology uses waveguides that are arranged at regular intervals on a square cartesian grid [12]. Figures 2.23(a) and 2.23(b) illustrate the regular 4-port scattering junction, and its use in a

mesh of arbitrary shape, respectively.

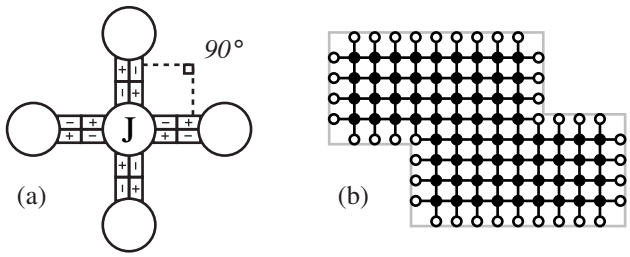


Figure 2.23: *Rectilinear topology: (a) the 4-port junction and (b) arbitrary shape mesh*

Boundary junctions, indicated by white circles in Figure 2.23(b) are implemented in 1, 2 or 3-port form, depending on edge orientation. Each regular scattering junction has four connections, evenly spaced at 90° from one another. The scattering equation for such a junction is

$$p_J = \frac{1}{2}\sum_{i=1}^{4} p_{J,i}^{+} \tag{2.119}$$

Calculation of the dispersion error using Von Neumann analysis [63] provides a measure of the accuracy of wave propagation simulation in the mesh model. The dispersion factor k is expressed as a function of two spatial frequency coordinates ξ_1 and ξ_2. It represents the ratio of actual mesh propagation speed c' to desired propagation speed c, in directional and frequency terms [64]. The centre point $k(0,0)$ of the resulting 2D plot is equivalent to DC. Any point on a circle of radius ξ away from the DC centre denotes the actual spatial frequency $\xi = \sqrt{\xi_1^2 + \xi_2^2}$. Temporal frequency is determined by $f = c\xi$. It can be shown that k can be derived as [65]

$$k(\xi_1,\xi_2) = \frac{c'(\xi_1,\xi_2)}{c} = \frac{\sqrt{2}}{2\pi\xi}\arctan\left(\frac{\sqrt{4-b(\xi_1,\xi_2)^2}}{b(\xi_1,\xi_2)}\right) \tag{2.120}$$

Where $b(\xi_1,\xi_2)$ is a geometric factor related to the orientation of connections in the topology with $\omega_1 = 2\pi\xi_1$ and $\omega_2 = 2\pi\xi_2$ [65]. For the rectilinear mesh b

is

$$b(\xi_1, \xi_2) = \frac{1}{2}\left(e^{j\omega_1 cT} + e^{j\omega_2 cT} + e^{-j\omega_1 cT} + e^{-j\omega_2 cT}\right) = \cos(\omega_1 cT) + \cos(\omega_2 cT) \tag{2.121}$$

Figure 2.24 demonstrates the dispersion factor for the rectilinear mesh.

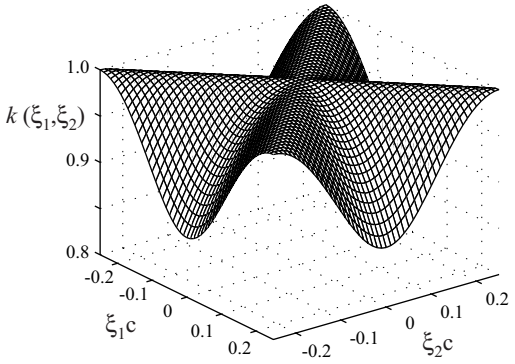


Figure 2.24: *Rectilinear DWM dispersion factor*

Ideal propagation, such that signals travels at the same speed, independent of frequency content or direction would show a flat dispersion factor such that $k(\xi_1, \xi_2) = 1$, as observed at the DC centre. Directional scattering properties are depicted in terms of an angular rotation about the centre, in accordance with the spatial frequency coordinates. Here, the two waveguide axes in the rectilinear mesh run parallel with ξ_1 and ξ_2. Figure 2.24 shows the ideal propagation speed for all frequency content along the diagonal of the rectilinear mesh where the dispersion factor is unity. In the axial direction, a reduced wave speed is experienced by higher frequency components.

Nyquist theorem states that in a sampled system, frequency content at up to half the sample rate can be sustained. A further limitation of the rectilinear mesh is that the resulting output is only valid up to $f_s/4$ [58]. This is because of the number of waveguides in any available pathway between two junctions will always be odd or even, but never a combination. For example, a pathway can be traced between two neighbouring junctions through $1, 3, 5, 7, \ldots$ waveguides. Consider the input of a unit impulse to a junction on a mesh where all surrounding nodes are at zero pressure at

discrete time (n). At $(n+1)$, the scattering equation (2.119) dictates that surrounding junctions at one waveguide away will receive some scattered pressure, and the central input junction will return to zero. By $(n+2)$ the wave will have spread to junctions at two waveguides away, but those at one waveguide distance will return to zero. The central junction will have been affected by its neighbours at this second time step and so will be non-zero. This chess-board effect means that any two junctions at one waveguide apart function alternately from one another [9]. The resulting effect is that the mesh output valid bandwidth is halved from $f_s/2$ to $f_s/4$, and any spectra produced will be mirrored about this point.

The Triangular Mesh

A triangular DWM is formed when the propagating medium is sampled such that a radiating circular wavefront is spread amongst 6 equally spaced connections [64]. Figures 2.25(a) and 2.25(b) depict the regular 6-port scattering junction, and its use in a mesh of arbitrary shape, respectively.

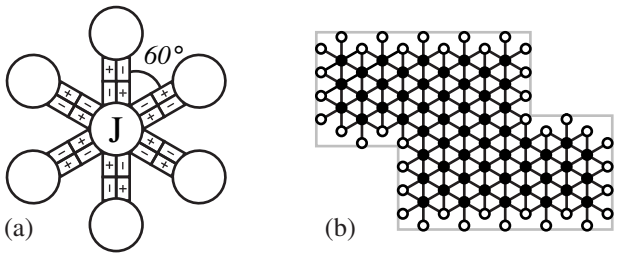


Figure 2.25: *Triangular topology: (a) the 6-port junction and (b) arbitrary shape mesh*

Boundary junctions, indicated as white circles on the diagram, are present as $1-5$ connection terminations. Each of the six connections at 60° from one another on a regular junction contribute to the scattering equation in the following manner.

$$p_J = \frac{1}{3}\sum_{i=1}^{6} p_{J,i}^{+} \tag{2.122}$$

This junction arrangement results in wave propagation that is more accurate than the rectilinear topology. The geometric factor b used in the calculation

of the dispersion error for the triangular mesh is

$$b(\xi_1,\xi_2) = \frac{2}{3}\left[\cos(\omega_1 cT) + \cos\left(\omega_1 cT/2 + \sqrt{3}\omega_2 cT/2\right) + \cos\left(\omega_1 cT/2 - \sqrt{3}\omega_2 cT/2\right)\right] \tag{2.123}$$

Figure 2.26 shows the dispersion factor for the triangular DWM.

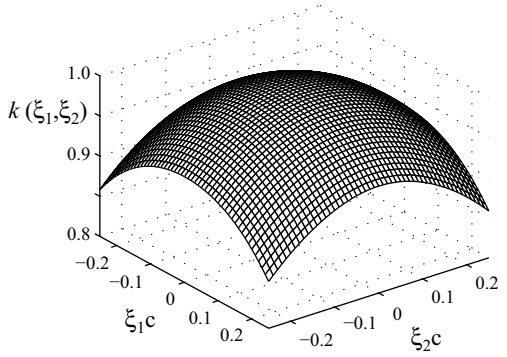


Figure 2.26: *Triangular DWM dispersion factor*

As with the rectilinear mesh, a reduced wave speed is experienced by higher frequency content. However, the graph bears greater circular symmetry about the centre. The triangular topology has less angular variation in terms of dispersion error, and therefore has a more even directional scattering. Furthermore, the triangular dispersion factor about the majority of the middle is approximately flat. This represents accurate propagation at frequencies that are relatively low compared to the sampling rate. In simulations where a wide bandwidth is required it is possible to reduce the dispersion error in higher frequency content with the use of frequency warping [66]. A warped FIR filter is used for pre- and post-processing of signals into and out of the mesh to correct for shifts in the higher frequencies.

The Interpolated Mesh

Further improvements in scattering uniformity are offered with the use of the interpolated mesh. An arbitrary number of connections are considered at a rectilinear junction such that they form a circle around it. Those that fall in between grid axes are purely hypothetical. Interpolation is used to

extrapolate the effect of those additional connections onto actual connections [67]. Figures 2.27(a) and 2.27(b) show the formulation of an interpolated scattering junction and its use in a DWM of arbitrary shape, respectively.

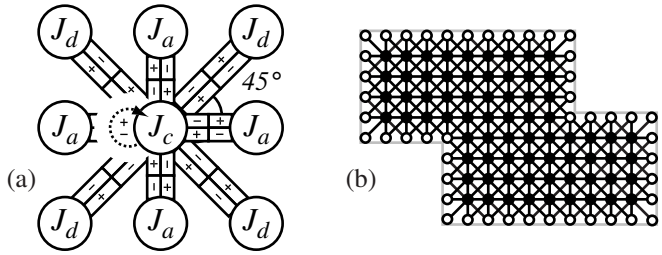


Figure 2.27: *Interpolated topology: (a) the 9-port junction and (b) arbitrary shape mesh*

A 3×3 scattering matrix is formed such that the influence of each of the 8 neighbouring junctions and the central junction itself, is defined. The weighting applied to the diagonal, axial and central components is h_d, h_a and h_c, respectively. Bilinear or quadratic interpolation, or alternatively an iteratively calculated optimum between the two methods, can be used to define the point spreading function [68].

$$h_{x,y} = \begin{bmatrix} h_d & h_a & h_d \\ h_a & h_c & h_a \\ h_d & h_a & h_d \end{bmatrix} = \begin{bmatrix} 0.09398 & 0.3120 & 0.09398 \\ 0.3120 & 0.3759 & 0.3120 \\ 0.09398 & 0.3120 & 0.09398 \end{bmatrix}_{optimum} \tag{2.124}$$

The pressure at the junction is then

$$p_{J_c} = \frac{2}{N} \sum_{x=1}^{3} \sum_{y=1}^{3} h_{x,y} p_{x,y}^{+} \tag{2.125}$$

The interpolated junction gives improvements in wavefront isotropy, but introduces an extra 9 multiplications. As such the triangular mesh is widely used in 2D modelling, considered to be accurate enough, whilst maintaining suitable computational simplicity [10] [64] [69] [58].

2.6.7 The Multiple-Port Finite Difference Junction

A finite difference scheme may also be applied to a multidimensional structure. As for the 1D case, PDEs with respect to time and space are approximated as second order differences. For example, the wave equation for pressure $p(x,y,t)$, in 2D Cartesian coordinates x and y is

$$\frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} = \frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} \tag{2.126}$$

Following the same method as in Section 2.4.3, a FDTD approximation can be constructed. Spatial coordinates x and y are indexed as m_x and m_y, respectively, and n is the discrete time step, such that the pressure p at a point is

$$\begin{aligned} p(m_x,m_y,n) &= p(m_x,m_y+1,n-1) + p(m_x+1,m_y,n-1) \\ &+ p(m_x,m_y-1,n-1) + p(m_x-1,m_y,n-1) \\ &- p(m_x,m_y,n-2) \end{aligned} \tag{2.127}$$

A general form of the scattering equation for pressure p at a FDTD junction J with N ports of admittance Y can be derived [54] as

$$p_J(n) = 2\frac{\sum_{i=1}^{N} Y_i p_i(n-1)}{\sum_{i=1}^{N} Y_i} - p_J(n-2) \tag{2.128}$$

For a junction with N equal admittance connections, this becomes

$$p_J(n) = \frac{2}{N}\sum_{i=1}^{N} p_i(n-1) - p_J(n-2) \tag{2.129}$$

2.6.8 Finite Difference Boundary Implementation

The absence of travelling wave variables makes the FDTD scheme inflexible to the formulation of multiple-port boundaries. The loss of the directional information of incoming components at such a junction means that directional weighting of outgoing components is also lost. A 2D mesh using only FDTD scattering is therefore limited in shape to a rectangular, rectilinear grid

with edges that are parallel to the two cartesian coordinate planes x and y. Boundaries are therefore straight and can be implemented using a line of 1D terminations. Simple 1-port reflecting FDTD mesh boundaries can be simulated using the same admittance discontinuity analysis techniques used for the wave scattering boundary in Section 2.5.3. The pressure p on a 1-connection FDTD boundary junction B in terms of the singular neighbouring junction J is

$$p_B(n) = (1+r)p_J(n-1) - p_B(n-2) \tag{2.130}$$

This boundary formulation is derived assuming 1D wave motion. When used at the edges of a 2D or 3D mesh it only applies the desired reflection to the components on the wavefront that are perpendicular to the boundary at the point of incidence. All other components experience some small reflection. This effect is negligible in the case for r values approaching fully positive or fully negative reflections. However, it becomes apparent in the case of an absorbing boundary, where $r \approx 0$. A Taylor series approximation to the pressure leading up to the edge of a mesh can be used as an improved absorbing boundary condition [70]. Further advancements towards directionally balanced 2D FDTD boundary conditions have also been achieved using a stepped impedance layer at the edge of the mesh [71]. Boundary junctions employ a spatial averaging filter to take into account the effect of scattering junctions within this additional layer [72]. The spatial filter includes junctions on a line perpendicular to the boundary, as in (2.130) and [70], but also considers the influence of junctions to either side. This inclusion of surrounding pressure values from a broader range of incident angles improves the directional behaviour of the boundary.

2.6.9 Wave Scattering vs FDTD: Mixed Modelling

The FDTD junction consists of three variables for any number of connections in any dimensionality or topology. These are the total junction pressure values at the three time instants (n), $(n-1)$ and $(n-2)$. For each wave-scattering junction a total pressure value is required with two travelling components for

each connection. For a 1D system, wave-scattering methodology is preferred as it offers the potential of combined delay lines for one computation per time sample [10]. In higher dimensions, or where junctions have more than two connections, the use of FDTD is advantageous. Scattering equations contain fewer variables and operations per junction than wave-scattering. This results in signal processing algorithms that are more computationally efficient, placing lower memory and speed demands on a system. One of the main disadvantages of the FDTD scheme, however, is numerical instability. This is caused by rounding errors arising from the finite difference approximation to the differential operators in the wave equation. Wave-based scattering methodology is exact and is therefore numerical robust [73]. Furthermore, the geometrical inflexibility of FD boundary junctions places limitations on allowed mesh shape.

Clearly, both approaches to spatial discretisation offer different advantageous scattering properties; that of speed and efficiency in FDTD, and stability and geometric flexibility in wave scattering. The two scattering methodologies that have been presented here were seen to be equivalent in Section 2.4.4. It is possible to combine the two modelling paradigms to exploit the benefits of both with the use of the KW interface [54].

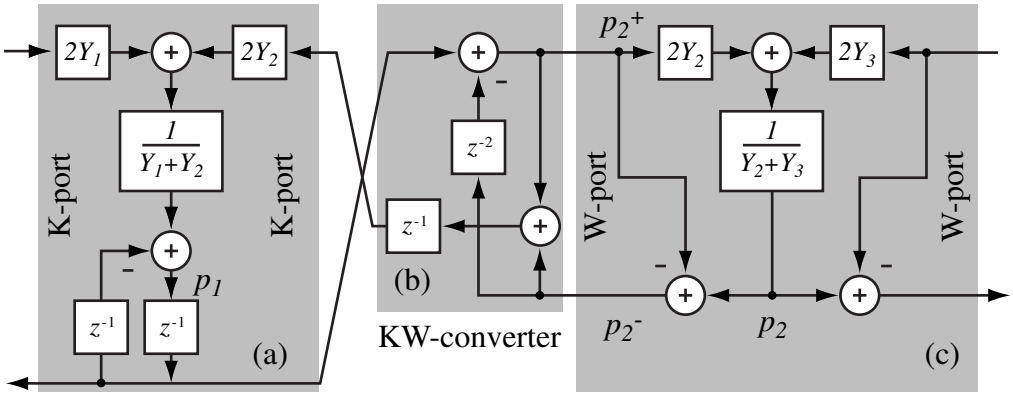


Figure 2.28: *Signal processing schematics for the (a) K-node, (b) KW-Converter and (c) W-node, after [54]*

These units provide a stable interface between Kirchoff (K) and Wave (W) variables. K scattering is obtained with a linear transformation from W-

variables, replacing travelling wave variables with past pressure values from neighbouring junctions. This is an alternative approach to FDTD for use in mixed modelling although there does not always exist an exact equivalence between an FDTD model and linearly transformed W-variable K-model, for instance the lack of appropriate equivalent boundary solutions.

Using mixed modelling, a multidimensional simulation of, for example, the acoustics of a concert hall can be constructed that presents improvements in computational efficiency [15]. The majority of the waveguide structure away from the boundaries that consists of standard N-port scattering junctions can be simulated with K methodology as to benefit from the reduced computational load. The geometrical flexibility advantages of the wave-scattering method at the boundaries can also be exploited. Furthermore, with wave based N-port boundaries, the K scheme can also be implemented as a different topology mesh as to benefit from the improved dispersion characteristics.

2.7 Conclusions

In this chapter the use of digital waveguides in an acoustical physical model has been discussed. Fundamental physical quantities, such as the speed of sound in air, and the acoustic impedance experienced by a propagating sound wave, were derived from first principles. The natural phenomenon of resonance was introduced in order to highlight how wave reflections within a confined system give rise to modal frequency peaks in the resulting spectrum. These concepts of sounds waves and resonance are of particular interest in speech studies. As will be seen in the next chapter, the human voice is an acoustic resonator which is continually manipulated by its user to alter the way in which a sound wave propagates through it. This process changes the resonant qualities of the vibrating air-cavity within, giving rise to the many different voice sounds we interpret as spoken communication.

Next, the numerical simulation of a real world system was discussed in two general stages. Firstly, identification of a continuous wave equation

governing system behaviour takes place. Second, a solution is found that satisfies the wave equation within a given discretisation of the system. One of the methods presented, the travelling wave solution, forms the underlying theory of the digital waveguide. The manner in which this modelling paradigm is used to construct a 1D simulation was examined through the formulation of various scattering equations. This lead to the derivation of the Kelly-Lochbaum concatenated tube model, which was originally presented in 1962 as a 1D model of the vocal tract.

Finally, the use of the waveguide in a model with higher dimensional representation was shown. The digital waveguide mesh was defined in terms of multiple port scattering and boundary junctions. These techniques are used in acoustical simulations of 2D surfaces, such as drum skins, and 3D structures such as the acoustics of a room. 2D simulations are also often used as an efficient precursor to full 3D systems, offering a proof-of-principle examination of new techniques, such as boundary junctions, without adding full complexities too early in the development stage. The analysis given in this chapter demonstrating the expansion of the 1D waveguide model into a multi-dimensional DWM is fundamental to the work contained in this thesis. It lays down the process by which it will be suggested in the following chapters that the well-established 1D vocal model can be extended into a multidimensional model.

In all simulations, a number of assumptions and simplifications have to be made to the definition of the synthesis system. The model needs to capture enough essential aspects of the real world behaviour to meet the appropriate level of representation that is required. At the same time, it is necessary to determine real-world properties that are of no interest or consequence to the desired performance, such that they can be omitted. Similarly, the modelling methods themselves are not exact. The nature of discretisation implies that some properties of the representation will be lost. Some of the deficiencies of the modelling methods were discussed in this chapter, such as the linearisation applied in the finite difference approximation of the wave equation, or the dispersion characteristics of each of the different topology

DWMs. These issues will be present in constructing a multidimensional DWM model of the vocal tract, and are worth consideration. However, as will be outlined in the next chapter, the human voice is a very variable system. Large amounts of uncertainties will be present in voice modelling techniques, and so achieving a perfect vocal synthesis system is not an expectation at this stage in the research.

Chapter 3

The Human Voice

3.1 Introduction

The human voice is a highly complex system that has evolved over many years to allow communication between humans using a wide variety of speech sounds. We use the lungs and vocal folds to create pressure waves that undergo propagation and reflection within the resonating cavity formed by the vocal tract. These acoustical disturbances are manipulated into spoken communication using only a few articulatory movements. Speech synthesis provides an alternative whenever human speech is not possible or practical. In the most widely known application it functions as an artificial voice for the vocally impaired. It also serves as an aid for the visually impaired in conveying the content of an electronic document or email. In the broader context of human-computer interaction, many further situations exist where a computer system is required to audibly communicate with its user.

Natural sounding speech synthesis, such that it is indistinguishable from a human speaker, is the goal in state-of-the-art applications. However, the highly variable nature of speech, and complexity of the vocal system used to create it, make this a non-trivial task. Recent developments in techniques based on the joining together of short extracts of recorded speech have resulted in artificial speech of a highly organic nature [6] [22] [5]. Ultimately, however, such sample-based methods are limited to reproduction of only

the sounds that were originally recorded. Articulatory models mimic the actions of the vocal system, rather than reconstruct the resultant sound. The high levels of complexity in such a model make them impractical for current requirements. Low-level articulatory models that make many simplifications to the vocal process, are widely used in speech research. Continual advances in knowledge about the voice, alongside a constant increase in available computational power, have stimulated interest into higher levels of representation. At current levels of sophistication, natural articulatory speech synthesis is largely a future consideration. However, research into techniques to improve the accuracy of simulations of the constituent vocal sub-systems is widespread.

This chapter starts off by briefly looking at the human vocal anatomy and its function as an acoustic resonator. The vocal tract, and the manner in which it is used to create the many speech sounds, is examined. Vocal fold vibrations and the resulting glottal waveform are introduced. Phonetic descriptions used follow the IPA notation [74]. For example the vowel in the word *bed* is identified with the symbol /ɛ/. A table of IPA descriptions of vowels, diphthongs and voiced and voiceless consonants, along with relevant word usage can be found in Appendix B.1. Next, three different methods of generating artificial speech sounds are outlined:

- Formant synthesis - Reconstruction of the known spectral properties
- Concatenative synthesis - Joining together pre-recorded samples
- Articulatory synthesis - A model that mimics the physical vocal process

Particular focus is directed towards the time-domain acoustic tube articulatory vocal tract analogy. A thorough analysis is presented of the widely established 1D Kelly-Lochbaum system. Finally, the potential of extending the dimensionality in such a model is highlighted.

3.2 Acoustics of Voice Production

The term *vocal tract* is used to describe the air cavity that is used in production of human speech. It comprises three resonating chambers. Between the vocal folds and the lips are the *pharyngeal cavity* followed by the *oral cavity*. The *nasal cavity* forms a side branch from the oral cavity at the velum and extends to the opening at the nostrils. The term vocal tract also includes the various features that surround the cavities, such as the larynx, epiglottis, pharynx, tongue, soft palate (or velum), hard palate, teeth and lips. Figure 3.1 illustrates these acoustically important features of the human head from a sideways-on cross-sectional view, known as the *mid-sagittal* plane.

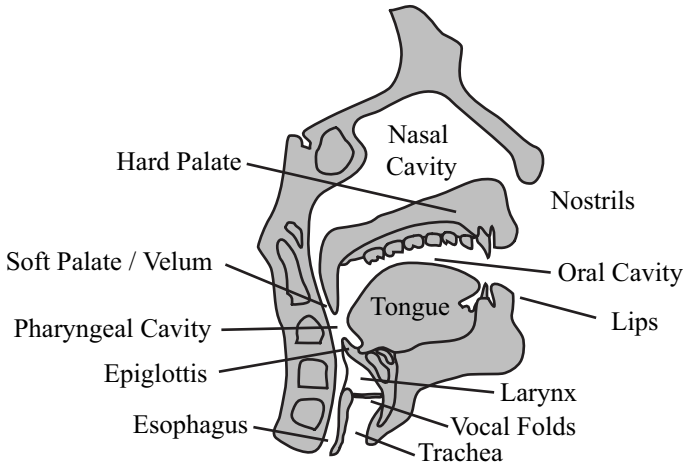


Figure 3.1: *The human vocal system, after [75]*

The sounds created during continuous speech can be considered as the result of a source-filter combination [76]. The vocal tract serves as a variable shape acoustic filter that is continually manipulated by the speaker in order to change its resonant properties. Excitation comes in the form of a stream of pressure pulses that are generated as air flowing up from the lungs causes the vocal folds to open and close periodically. The vocal tract filter acts to impart its spectral characteristics onto the glottal source signal, such that the resulting output from the lips is perceived as speech sounds.

3.2.1 Vocal Tract Area Function

The full vocal tract shape can be acquired using either x-ray [76], magnetic resonance imaging (MRI) [77] or acoustic pulse reflectometry [78]. For each vowel shape, the subject will be asked to hold a tract position while the scan takes place. During continuous speech the tract shape is continually changing. Data acquired from a scan of a stationary tract position held for some time will not, therefore, be an exact representation of that which would be observed in speech. Static tract shapes are adequate for current levels of speech modelling. However, methods introducing higher levels of accuracy will be of benefit to future research.

For simplicity in modelling purposes the tract shape is generally quantified in a simplified 1D *area function*. Figures 3.2(a) and 3.2(c) illustrate the tract in the position held for production of the /i/ vowel as in the word *bead*, and the resulting 1D area function, respectively. The velum is closed for production of such non-nasalised vowels and so the nasal cavity is not included in the diagram. Figure 3.2(b) shows how the 1D data is extracted, taking a plane that is roughly perpendicular to the localised airflow for the orientation of the cross sectional area. This process has the effect of straightening the tract.

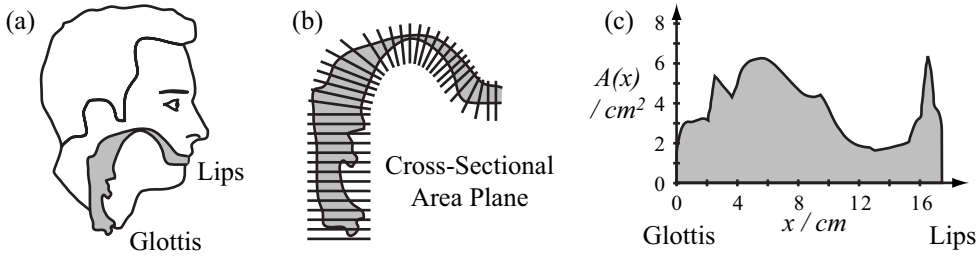


Figure 3.2: *The vocal tract in the /i/ vowel position: (a) in situ, (b) area plane for cross-sectional orientation and (c) resulting 1D area function*

Information about the bend in the tract and the actual cross-sectional shape is lost in this process. The 1D area function therefore provides an approximation in the form of a series of circular area values projected along a straight axis from the glottis to the lips.

3.2.2 Nasal Tract Area Function

The soft palate or velum is used to control the air-flow into the nasal cavity. For velar opening areas of 20 mm^2 or less the sound produced is non-nasal [79]. Acoustical coupling between the vocal and nasal tracts takes place when the opening is in the region of 50 mm^2 [80]. The sinus cavities also contribute to the nasal resonances. The combined nasal and sinus cavities form an additional resonator of approximately 11 cm in length. The cross sectional width variations can be quantified in an area function. Figure 3.3 shows an example of the measured nasal tract area function.

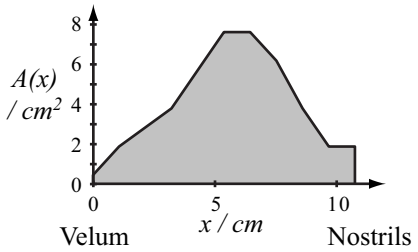


Figure 3.3: *The nasal tract area function, after [81]*

In general, nasal consonants are characterised by a formant at between 200-300 Hz in the speech spectrum [76]. During production of the nasal consonants /m/ and /n/, the vocal tract is closed and the sound radiates through the nostrils. This type of acoustical coupling generates antiresonances (inverted peaks) in the spectrum, the lowest of which are prominent at around 1000 Hz for /m/ and 1700 Hz for /n/ [82].

3.2.3 Glottal Excitation

The source of excitation to the vocal tract is a combination of the air pressure direct from the lungs and the vibrations of the vocal folds that results from this sustained pressure. Located in the larynx above the trachea, the vocal folds are two parallel mucosal membranes. They are attached at the front of the larynx to the thyroid cartilage and at the back to the arytenoid cartilages. The space in between the folds is called the *glottis*. For production of voiceless vowels (whispered) and consonants the glottis remains open.

In preparation for voiced phonation the arytenoids are made to rotate to bring the folds closer together, causing tension across them. A contraction of the diaphragm muscles forces a steady stream of air up from the lungs. The air pressure builds up behind the closed glottis. When the force against the vocal folds is greater than the elastic tension holding them together it causes them to move apart, briefly releasing a pulse of air. Bernoulli's principle of fluid dynamics states that this increase in flow velocity will occur with a decrease in pressure. The tension in the vocal folds, combined with the reduction in pressure, will cause the glottis to abruptly return to its closed state. This periodic cycle produces a train of pulses into the tract at an average fundamental frequency of 120 Hz for a male speaker [79].

Models of the glottal vibrations generally fall within two categories; those that describe the motion of the vocal folds [83] [84] [85] [86] and those based on a mathematical description of the resulting waveform [87] [88] [89].

Mass and Spring Models

The movement of the vocal folds during the periodic opening and closing can be modelled with a mass and spring representation. Each fold can be viewed of as a singular mass which moves in to meet the opposite fold in the centre of the glottis. The spring represents tissue stiffness, or restoring force provided by the muscles. Some damping may be included to represent the energy absorption of the tissue. This simplistic model assumes that the whole of the fold moves as one rigid mass and excludes effects of the flexibility of the membrane on the resulting waveform. More accurate models have been developed to give greater degrees of freedom to the moving parts in each fold. These have been of the form of two- [83], three- [84] and sixteen-mass [85] models. However, increased flexibility and accuracy resulting from these models is offset with the greater computational complexity needed to facilitate them.

The LF Glottal Waveform Model

The Liljencrants-Fant (LF) four-parameter glottal flow derivative model [89] is commonly used as excitation in vocal simulations. It provides a succinct mathematical description of the air flow through the glottis as it opens and closes. Figures 3.4(a) and 3.4(b) illustrate one cycle of the glottal flow $u(t)$ and differentiated glottal flow $g(t) = \frac{du}{dt}$ waveforms, respectively.

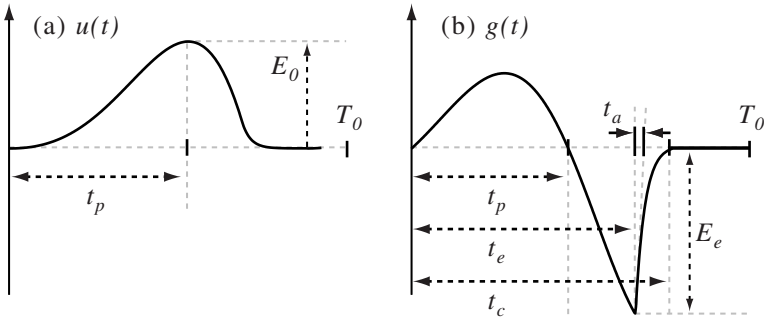


Figure 3.4: *Glottal waveforms: (a) flow and (b) flow derivative*

The four timing parameters needed to generate the waveform are t_p, t_e, t_c, and t_a. Different combinations of these values can be obtained from natural speech of various voice type, such as modal (regular voiced phonation), breathy or whispered. The glottal waveform for the associated voice type can be reconstructed using the parameters. Time t_p occurs at the moment of maximum flow E_0, as in Figure 3.4(a). Mathematically, the derivative waveform $g(t)$ in Figure 3.4(b) is defined in two parts. In the first section, an exponentially growing sinusoidal component is used to represent the time from glottal opening to its maximal negative value E_e, at time t_e. This sinusoid has angular frequency $w_g = \frac{\pi}{t_p}$. Secondly, the residual flow after t_e is represented with an exponential component as the waveform returns to zero at t_c. Effective duration of the return phase is given by t_a. The cycle is complete at T_0. The two stages of the waveform are

$$g(t) = E_0 e^{\alpha t} sin(\omega_g t) \qquad 0 \leq t \leq t_e \tag{3.1}$$

$$g(t) = -\frac{E_e}{\varepsilon t_a}\left[e^{-\varepsilon(t-t_e)} - e^{\varepsilon(t_c-t_e)}\right] \qquad t_e \leq t \leq t_c \leq T_0 \tag{3.2}$$

Where α, the exponential growth factor of the sinusoid, is typically around 0.7, although is generally in the range $0.5 - 0.9$ [90]. The symbol ε denotes the exponential time constant of the return phase. The following conditions must hold.

$$\int_0^{T_0} g(t)dt = 0 \tag{3.3}$$

$$\varepsilon = \frac{1 - e^{-\varepsilon(t_c - t_e)}}{t_a} \tag{3.4}$$

$$E_0 = -\frac{E_e}{e^{\alpha t_e}\sin(\omega_g t_e)} \tag{3.5}$$

These constraints are established as to ensure that the model maintains continuity. The area balance expressed in condition (3.3) is achieved with iterative calculation of (3.5). During this, E_0 and α are modified to determine an equal distribution of positive and negative parts of the integral of $g(t)$ [91]. For small t_a values, the approximation $\varepsilon = 1/t_a$ may be made in place of (3.4) [89].

3.2.4 Vowel Formant Frequencies and Bandwidths

In simplistic terms, the vocal tract exhibits resonant behaviour that is analogous to a straight acoustic tube, such as that illustrated in Figure 2.8, with an open end at the lips, and a closed end at the glottis. This comparison is particularly appropriate in the case of the tract shape held for producing the neutral vowel /ə/, heard for example at the beginning of the word *about*. In this configuration the tract has the least cross-sectional variation, and so best approximates the straight tube analogy. The average adult male vocal tract measures 17.6 cm [76]. In Section 2.3.5 the modal frequencies along an acoustic tube with one closed and one open end were analysed. Using a wavespeed of $c = 343$ ms^{-1}, the first three modes of resonance ($N = 0, 1, 2$) along a tube of this length can be calculated with equation (2.38) to be 487 Hz, 1462 Hz and 2436 Hz. In the context of speech these resonant peaks are called *formants*.

The different vowels are created as the tract shape is moved away from

the neutral position, providing constrictions to the air-flow and changing the resonant properties. This has the effect of moving the formant frequencies, giving each vowel its identifiable characteristics. Figures 3.5(a) and 3.5(b) show the tract shape in the position held for the /ɪ/ and /ɔ/ vowels, respectively. The associated formant patterns, showing the first three peaks, are included in the diagram.

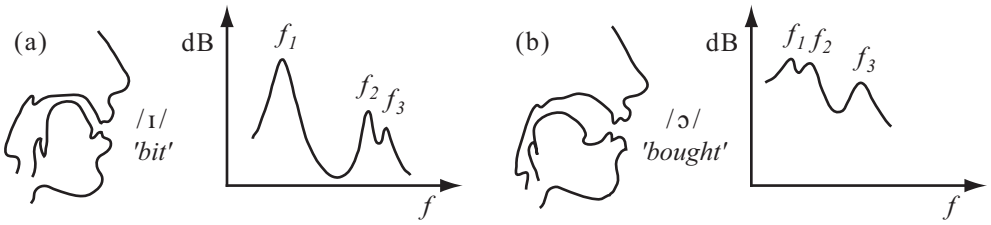


Figure 3.5: *Tract shapes and formant patterns: (a) /ɪ/ and (b) /ɔ/ vowels*

The diagram indicates that particular tract sections have greater influence over individual formants. For example the tract configuration in Figure 3.5(a) comprises a constriction made towards the front with the tongue, combined with a larger cavity towards the back. This type of tract shape results in a lowered f_1 and raised f_2 and f_3, when compared to their neutral positions [41]. Similarly, a constriction made further back, as in Figure 3.5(b), leads to a raised f_1 and lowered f_2 and f_3.

Figure 3.6 illustrates average formant frequencies and corresponding bandwidths (in brackets) measured from a range of vowels occurring in natural speech.

Formant frequencies are used to identify one vowel from another. In perceptual terms, only the first three formants are required for differentiating one vowel from another. Higher formants are considered to contribute to the unique characteristics of the speaker [41]. The bandwidth of a formant is defined as the width, or frequency range at a point 3 dB less than the peak value. Formant bandwidths determine the extent to which the formant frequencies affect the output. In other words, they control the quality of the vowel sound. A tract with large bandwidth formants would add little to the spectrum of the input source, and so the resulting speech output

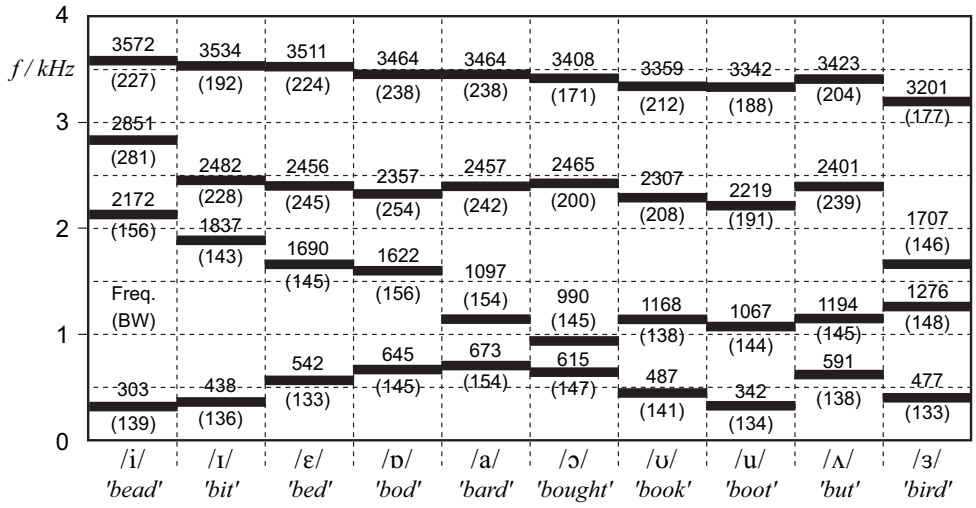


Figure 3.6: *Average formant frequencies and bandwidths for male speakers, after [79]*

would be perceived as similar to the glottal *buzz*. In general, increased losses in a system imply less inward reflections and therefore weaker resonant characteristics. In the opposite case, higher reflections result in lower bandwidths and a stronger resonance. Vowel sound that is produced using a system with overly narrow bandwidth formants is often described as having an unnatural ringing, metallic quality.

3.2.5 Energy Losses

The vocal tract air cavity is enclosed by three main boundaries. Reflection and absorption takes place at the vocal folds, tract inner walls and lip opening.

The Glottal Boundary

The energy absorption of the vocal fold tissue follows a frequency-dependent relationship with pressure waves incident upon it. However, on average approximately only 3.1% propagates through to the trachea [92], [93]. Neglecting vibratory effects it is therefore possible to consider the glottis as a positive (phase preserving) reflection of $r_g = 0.969$.

Vocal Tract Walls

The soft tissue inside walls of the tract also present frequency dependent positive reflections. Energy is absorbed as the yielding walls vibrate. Such losses are more pronounced at the lower end of the spectrum as little resonance of the large wall structure takes place at higher frequencies [94]. Friction between the air and tract walls, and heat conduction into the tract walls generates viscous and thermal losses. The effects of heat and friction losses are negligible at low frequencies [94]. A circuit-based analogy can be constructed such that it models the transmission properties of a section of acoustic tube, incorporating the effects of losses that would be observed in the vocal tract [83], [95], [96]. An example of this type of model is shown in Figure 3.7, which uses the analogue filter circuit components: resistor R, capacitor C and inductor L.

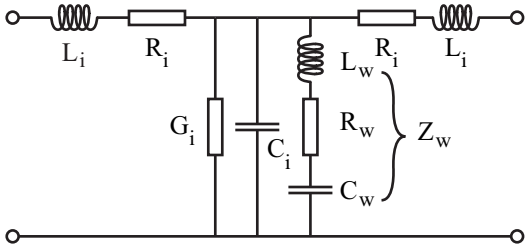


Figure 3.7: *Circuit-based vocal tract acoustic tube section analogy, after [79]*

Component	Equation	Acoustic representation
R_i	$\frac{2\mu l_i(a_i^2+b_i^2)}{\pi a_i^3 b_i^3}$	Viscous loss
C_i	$\frac{A_i l_i}{\rho c^2}$	Air compressibility
L_i	$\frac{\rho l_i}{2A_i}$	Air mass inertia
R_w	$\frac{B_w}{l_i S_i}$	Part of yielding wall
C_w	$\frac{l_i S_i}{K_w}$	Part of yielding wall
L_w	$\frac{M_w}{l_i S_i}$	Part of yielding wall
G_i		Heat conduction loss

Table 3.1: *Relationship between circuit components and acoustical quantities, after [79]*

Lip Radiation

The boundary between the vocal tract and the air in front of the speaker at the lips forms a negative (phase inverting) frequency-dependent reflection. An analogy is formed such that the lip opening is modelled as a piston at the inner edge of a spherical baffle (the head) that is forcing the air column in front of it to vibrate. If it is assumed that the radiating surface is small compared to the head, then the curvature of the sphere can be neglected. Using this simplification, lip radiation is commonly represented as an opening in an infinite plane baffle. A model using electronic circuit components results in the following load applied across the lip termination [94].

$$Z_l(\omega) = \frac{j\omega L_r R_r}{R_r + j\omega L_r} \tag{3.6}$$

This can be realised with a parallel resistor and inductor pair to represent a lip opening of radius a, where $R_r = 128/9\pi^2$ and $L_r = 8a/3\pi c$ and c is the speed of sound [96].

The lip opening has the effect of reflecting a greater proportion of the lower frequencies back into the tract. A further simplified model, which is more commonly used, implements this effect as a first-order high-pass operation such that the radiated pressure wave can be obtained with differentiation of the volume velocity signal [97].

3.2.6 Articulation

The tract shape is largely influenced by the arrangement of the articulators - the tongue, jaw, teeth and lips. Articulators are held in specific positions for production of voiced and voiceless (whispered) vowels. A slow movement of these features generates a transition between two vowels, called a diphthong. An example of this is /aʊ/, heard in the word *house*.

Sharp and abrupt articulator movements are used to create consonant sounds. At a point of constriction turbulence is generated in the air-flow. This gives rise to noise-like excitation called *frication*. The glottal approximant /h/, for example, is made with turbulence at the partially open vocal folds.

A constriction made with the lips and teeth produces the *labiodental fricative* voiceless /f/ or voiced /v/. The tongue and hard palate are used together to create the *alveolar fricative* such as /s/.

An obstruction to the airflow produces a complete stop. Upon release this produces a sudden impulsive excitation, used in the generation of *plosive* consonants. For example the sound /k/ where a stop is made between the tongue and velum. The *bilabial plosive* voiced /b/ and voiceless /p/ are made with a stop created with both lips. The tongue and hard palate are used for production of the *alveolar plosive* /t/. Slight constrictions can be made to the tract flow to make approximant consonants, where a small amount of turbulence is generated. For example, *semivowels* such as /w/ are made with the lips, and laterals such as /l/ are created with the tongue.

3.3 Vocal Synthesis

Advancements in speech technology research improve the way in which a human can interact with a computer in the situation where traditional interface devices such as a keyboard are not used. Direct spoken communication with a computer involves the process of *speech recognition*. In the reverse direction, the computer must provide audio output that is perceived to be speech and the meaning of the intended message must be correctly conveyed to the listener. This is *speech synthesis*. It can be broken down into two parts. Firstly, analysis and phonetic description. The word, message or concept intended for communication is expressed as a series of phonemes which still contain the original meaning. Secondly, audible output is generated that consists of the connected chain of phonemes, and hence is recognisable as the intended communication. It is this secondary stage of acoustic waveform generation of speech sounds which is of interest to this work.

3.3.1 Formant Reconstruction

Techniques based on spectral reconstruction make use of the known frequency characteristics of speech to create a system of filters. The formants

generated by the tract resonances can be modelled with a series of a notch filters with independently tunable frequency and bandwidth [88], [98]. These are continually modified to produce the formant patterns required for the range of different vowel sounds. Small pulses are injected into the system to model plosive consonants [99]. Noise based excitation can be used to produce frication and voiceless phonation. A glottal pulse train is injected into the system to provide voiced phonation. The system of filters imparts the spectral characteristics of the chain of phonemes to be synthesised onto the various excitations.

3.3.2 Speech Sample Concatenation

At present the most widely used method of speech synthesis uses concatenation of stored waveform samples. Several hours of spoken voice, typically that of an actor, are recorded and analysed. An instance of each of the possible units of speech, the phoneme, is extracted and stored in memory.

In what is known as text-to-speech (TTS) synthesis [6], the word, concept or message that is to be expressed as artificial speech will be input by a human or taken from an electronic document. A series of phonetic sounds will be selected from the database that provides a best-fit match to the desired utterance. The phonemes are connected together in a manner that minimises any perceptual cues that might reduce the resulting naturalness. Synthesised speech of a highly natural quality can be achieved with the use of unit selection [100]. Typically, a large diphone database database would be constructed, containing an example of each of the possible transitions from the middle of one phoneme to another [5]. At the concatenation stage context dependant selection is used to minimise the associated join cost function. This results in an increased match between a waveform and its adjoining segments and hence greater naturalness.

Concatenative synthesis does have its disadvantages. The database needed to store all the pre-recorded segments will become very large as the reqiurements of the system grows [22]. Advanced synthesis research involves

consideration of large vocabulary capabilities, alternative language support and emotional expression. Furthermore, the vocal identity that is provided with the system will alway be limited to that which was originally recorded as processing of the samples yields unnatural results [101].

3.3.3 Articulatory Modelling

Speech sound can be generated from a model that attempts to simulate the vocal system rather than reconstruct the resulting output. An articulatory model is a system in which moving tract parts are represented. Each is configured such that it provides functionality that approximates its observed role in the real world. In this sense it constitutes a physical model. Changes in tract shape can be directly applied to the controlling parameters of the model because of the semantic relation that exists between them. Furthermore, interpolation between tract shapes also has meaning, as interim parameters are physically realisable. In a spectral synthesiser interpolation between two sets of formants might yield patterns that correspond to unrealistic tract shapes.

The resulting signal that is actually simulated is the airflow as it is manipulated by the articulatory tract features. In most vocal tract models the two main assumptions made are that the tract is straight and that wave motion is one-dimensional. Simulations based on an acoustic tube with a straight axis, in part, allow for the simplified 1D representation to be justified. It has been demonstrated that within spectral regions of interest, the discrepancies that arise from this assumption are small [102]. A mathematical model of a curved duct in cylindrical coordinates was constructed to compare the difference in modal resonances with a straightened equivalent. For simplicity in calculations the duct was given a constant rectangular cross sectional area along its length. It was found that modal resonances below 4 kHz of a bent tube of constant cross sectional area differ from a straightened equivalent by $2\% - 8\%$.

Typically, in simulations of musical instruments the wave motion is also

assumed to be planar. The cross-axial modal resonances across a narrow acoustic bore are typically considered to be above the region of concern for standard audio applications. For example, the lowest cross-modal resonance of the narrow acoustic bore in a clarinet has been calculated to be 26.2 kHz [46].

In the frequency domain, an articulatory model consists of several controllable transfer functions. Sections of tract (that are not necessarily equal in size) are identified as important aspects within the speech system, such as the lips, nasal cavity, and tongue. The space contained within each section is used to derive a 2×2 or ABCD matrix [103]. Each represents the transfer function of the section. It takes into account effects of yielding walls, viscous losses, radiation and the position of articulators for the given vowel or area function. The combined product of the chain matrices gives the complete tract transfer function. A time domain glottal signal is convolved with the impulse response of this function to generate the speech like output. Improvements on this frequency domain approach make attempts to derive transfer functions for each section from 3D data, using the complete MRI scan, rather than a reduced 1D area function [101]. This incorporates some effects of non-circular tract segments on their propagational behaviour. However, the cascaded series of transfer function sections amounts to a 1D representation.

A visual simulation of the intricate movements of the articulators during speech has been constructed [104]. Full 3D MRI scans are taken of a speaker. The data is parameterised to produce a graphical model of the articulations, with particular emphasis on simulation of the complex shapes formed across the contours of the tongue.

Physical limitations can be applied to an articulatory model, for example, such that the tongue section can only move with the degrees of freedom observed in the real world. In this manner a comprehensive model could, to some degree of accuracy, achieve any of the sounds attainable with the human voice. If these limitations are extended beyond real world applications, then experimentation can take place that would otherwise not

have been possible. This could include synthesis of speech from a very small or very large vocal tract, such as that of a child, or a tall adult, respectively. Furthermore, the potential for obscure speech synthesis arises, such as a model that uses tract area functions taken from an animal.

3.4 The Time-Domain Acoustic Tube Vocal Tract Model

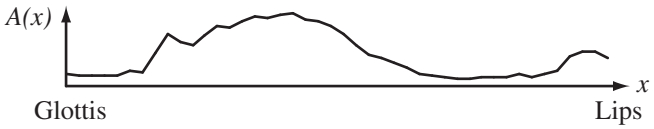
3.4.1 One-Dimensional Representation

The piecewise acoustic tube model is a widely established method of time-domain articulatory modelling of the human vocal tract. It was initially used to demonstrate the Kelly-Lochbaum scattering junction for signal distribution about an impedance discontinuity [18]. The travelling components solution to the wave equation is used to simulate the pressure signal in each equally sized tube section [94]. An in-depth analysis of the 1D digital waveguide and its application in a piecewise acoustic tube model was presented in Section 2.5.2. Developments of the model have been directed towards the use of wave filters (equivalent to resistor-capacitor-inductor small circuit tube approximations) in the model [105], [92], [106]. It has also been used to generate singing synthesis [107], [93].

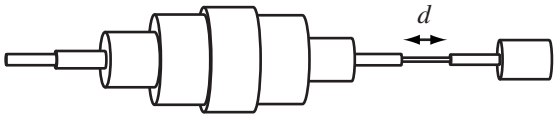
Given the assumptions made in straightening the tract and on planar wave motion, the vocal tract is represented as series of adjoining tubes. This type of model is called the 1D piecewise or concatenated acoustic tube model. The 1D area function is spatially sampled and represented as a number of discrete tubes of varying circular cross-sectional area. Figure 3.8(a) illustrates the area function for the /i/ vowel taken from MRI data [77]. Figures 3.8(b) and 3.8(c) show the associated piecewise analogies using equally sized cylindrical, and conical acoustic tube segments, respectively.

3.4.2 Conical Tube Sections

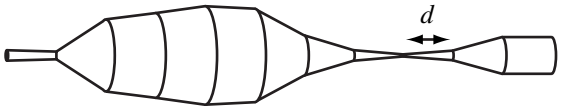
The spatially sampled cylindrical tube model shown in Figure 3.8(b) can be considered a zero-order approximation to the area function. The conical tube



(a) /i/ Vowel cross-sectional area function



(b) Concatenated 11-cylinder acoustic tube analogy



(c) Concatenated 11-cone acoustic tube analogy

Figure 3.8: *The 1D /i/ vowel waveguide model*

model illustrated in Figure 3.8(c) follows a first-order approximation. An improvement in the agreement of simulated formants to those predicted in theory is achieved with the use of conical waveguide segments [108]. As discussed in Section 2.5.4, additional filter units are formed within each junction to simulate the effects of spherical curvature of the wavefront as it passes through the cone. However, it has been demonstrated that this improvement comes with an increased computational load that is equivalent to a doubly spatially sampled cylindrical tube model, with no further gain in accuracy [109]. Therefore, with equal performance obtained from either model, the simplicity of the junction implementation in the cylindrical tube model makes it preferential.

3.4.3 The Nasal Tract

Nasalisation can be introduced into the 1D model [107]. The nasal tract area function is translated into a 1D piecewise acoustic tube model. The scattering junction at the nearest discrete point to the velum on the vocal tract model is configured with an additional port such that it provides a side branch opening into the nasal model. This third connection is the velum. The coupling between the two cavities is controlled with the admittance of the third waveguide.

3.4.4 Energy Losses

The output pressure at the lips is typically modelled with the application of a 6 db/octave high-pass filter to the velocity signal at the lips [107]. Reflections back into the tract at the lips are the remainder of this radiation function. Tract losses along the inner walls can be combined and projected onto either of the terminations. This is possible as the commutative properties of waveguide modelling allow for the separation of material properties from the purely propagational mechanism. The glottal boundary provides a high positive reflection, as indicated in Section 3.2.5.

Figure 3.9 demonstrates how the tube representations are modelled with a 1D chain of waveguides that are separated by junctions and terminated with glottal reflection r_g, and lip radiation filter $R_l(\omega)$ boundaries. Included in the diagram, the nasal cavity model is fixed at the velum and terminated at the nostrils with the radiation filter $R_n(\omega)$.

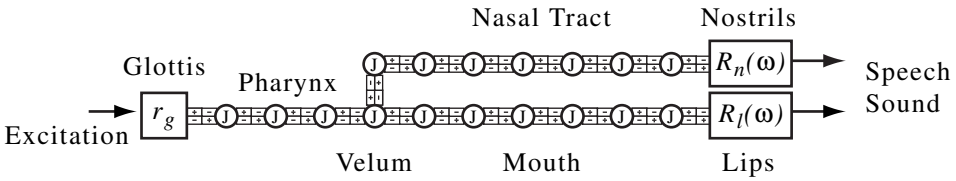


Figure 3.9: *The 1D waveguide vocal tract model with nasal cavity and radiation filters*

For the cylindrical model, the admittance of each waveguide is determined from the area of the associated tube with (2.11). Propagation through each waveguide junction J at the tube-area discontinuities is facilitated with a KL scattering unit, as discussed in Section 2.5.2.

3.4.5 Dynamic Operation

The crosswise tract shape changes that take place in speech are simulated with linear interpolation between area functions. This is implemented with continual adjustments in the admittance of each waveguide. One disadvantage of the 1D waveguide tract model is that the length must remain fixed. Production of certain vowels, such as the /u/ vowel, involve a rounding and protrusion of the lips. This has the effect of a slight lengthening

of the tract and hence lowering in formants. Dynamic alterations in the length have been simulated with the use of fractional delay waveguides in such a model [108], [110]. The additional filter units required introduce an increased computational load.

3.4.6 Computational Considerations

For illustrative purposes, Figure 3.9 shows a model using 11 waveguides to represent the length of the vocal tract. Typically a 44-waveguide model is used for sufficient accuracy in area function approximation. For a 17.6 cm vocal tract this gives a waveguide length of 0.4 cm, and a sampling frequency of $f_s \approx 88$ kHz. Using the one-multiply junctions for signal-flow at an impedance discontinuity, 44 multiplications and 176 additions are required per time-step. Including the few extra operations needed for boundary calculations, a real-time response is easily achieved on a standard PC.

3.4.7 Extended Dimensionality

A vocal tract model that employs a discretised 1D version of Webster's horn equation assumes that the wavelength of the speech waveform is much larger than the width of the tract. This implies that cross tract reflections are therefore small and occur at high frequencies so are beyond the scope of low bandwidth simulations. Minimal research has been undertaken to investigate the effects of wave motion in a tract model beyond planar considerations into a representation of higher dimensionality.

The reflections across a straight acoustic tube can be predicted using a Bessel function, as demonstrated in Section 2.3.5. It was shown that the lowest cross-modal behaviour of a clarinet with a radius of about 8 mm exists at frequencies of 12.56 kHz and 20.8 kHz. Such high frequencies are typically above the regions of concern for low bandwidth simulations, especially those above the limits of human hearing at about 20 kHz. For simulations of musical instruments this is often the justification for reduction

of the wave propagation mechanism into 1D planar considerations [46] [35]. Many vocal tract simulations use the same assumptions for computational simplicity, although equivalent radii of up to 20 mm are observed in speech. Using the same method (2.41) and the same two lowest zero-gradient values of $\alpha'_{11} = 1.841$ and $\alpha'_{21} = 3.054$ for the Bessel function with this larger radius, $a = 20$ mm, cross tract reflections occur at

- $f_{11} = \frac{343 \times 1.841}{0.02 \times 2\pi} = 5.03$ kHz
- $f_{21} = \frac{343 \times 3.054}{0.02 \times 2\pi} = 8.3$ kHz

In more simplistic terms, treating the vocal tract as rectangular duct with a constant width of 40 mm, the lower limit for cross modes calculated with the universal modal frequency equation (2.29) is 4.3 kHz. Tract openings of this size take place at only a few locations, such as in the mouth, and also occur infrequently in speech. However, these approximate lower bounds do serve to indicate that the modal resonances across the tract fall within the regions of interest of high bandwidth simulations.

3D finite element (FE) models of the vocal tract have been constructed to investigate the effects of higher order cross modes [19], [20], [111]. FE models present a similar method to finite difference (FD) for integration of partial differential equations. In [19], wave propagation in the 3D tract model was simulated with a spatial discretisation and calculation of frequency-domain interactions between the elements. Lip radiation was simulated as a piston in an infinite plane baffle [96]. Effects of yielding wall impedance were considered to be negligible above 1 kHz and so were omitted. The tract wall and glottal ends were implemented as hard reflecting boundaries. It was concluded that the multi-dimensional model demonstrated higher-order cross tract modes from 5 kHz upwards that were not accounted for in a 1D electrical circuit based representation [96] used for comparison.

A 2D time-domain transmission line matrix (TLM) model of the vocal tract has been used to demonstrate cross-modes [21]. Configured using the /a/ vowel area function, the authors report the presence of transverse modes at 5.2 kHz at the widest sections of the tract.

3.5 Conclusions

The acoustical characteristics of the human voice and ways in which it can be simulated have been discussed in this chapter. Emphasis was placed on production of voiced vowels and their measured area functions and associated formant patterns. The movement of the articulators was highlighted as the cause of manipulations to the air-flow which creates the variety of sounds observed in connected speech. Three methods for synthesis of the voice were presented. Filter-based formant reconstruction was briefly described. Sample-based concatenative methods were introduced as the most widely used technique, yielding the most natural sounding speech.

Physically-based articulatory models were identified as a method of vocal simulation with a potential for producing highly organic speech. Various models, and the simplifying assumptions made for each were outlined. It was indicated that a comprehensive vocal model of this type is currently an unrealistic goal. Despite this, physical models of some of its smaller subsystems, such as the mass-spring glottis and 3D tongue visualisation, were discussed to emphasise the potential of research in this area. It was noted that little has been done to investigate the effects of increased accuracy in the propagational subsystem of such a model. Some studies examining this notion in models were presented that reported cross-tract modal interactions at frequencies above 5 kHz. In order to continue these considerations, suggestions were made on extending the dimensionality of wave representation in the time-domain waveguide vocal tract acoustic tube model.

In general, this chapter demonstrates the largely variable nature of the voice, and some of the many efforts to understand, characterise and resynthesise aspects of it. It shows that physical modelling of the voice is a potentially useful field, into which several studies have been conducted, but also one in which there is still much to be discovered.

Chapter 4

The 2D Digital Waveguide Mesh Vocal Tract Model

4.1 Introduction

The widely established time-domain waveguide articulatory vocal tract model was examined in Section 3.4. The main aim of this project is to investigate the effects of increased dimensional representation as a direct extension to the traditional 1D acoustic tube model.

The pressure waves in the vocal tract can be modelled with a 2D representation, such that the assumptions on planar motion are removed. Simulation of the acoustic waveform as it is manipulated by the vocal tract articulators will therefore include propagational and reflectional pathways across and along the tract. The technique of extending the 1D waveguide into a 2D digital waveguide mesh (DWM) was examined in Section 2.6. The 2D DWM is applied here as a physical model of the resonator formed within the human vocal tract.

This chapter begins by demonstrating the differences in construction and frequency response between the 1D and 2D waveguide models. Two methods by which a cylindrical acoustic tube can be modelled with a 2D plane are discussed and one is selected for further examination. It is then considered how such a model can be tested. The manner in which the

vocal tract area functions can be applied to the width of mesh is discussed, and a brief description of the software used to run the simulations is also included. 2D DWM tract simulations are presented for a number of vowels, including /i/ in the word *'bead'*, /a/ in *'bard'* and /u/ in *'booed'*. These three are selected as between them they provide sufficient variation in articulator movement and formant frequency to examine the model under a broad range of attainable tract positions. The formant frequencies and bandwidths of simulated vowels are then analysed with reference to the 1D model. Lastly, problems associated with making dynamic shape changes to the 2D mesh model are discussed.

4.2 1D - 2D Comparison

To begin with, it is worth demonstrating the difference between the 1D and 2D physical modelling paradigms. Analysis of the construction and frequency response of a rectangular 2D DWM highlight the additional spectral content included with respect to a 1D waveguide of equivalent length.

4.2.1 1D Chain

A 17.6 cm straight tube model can be constructed using a 1D chain of 176 equal impedance digital waveguides, where each represents a 1 mm section. This results in a sampling frequency of $f_s = 343$ kHz. With full positive reflections ($r = 1$) at either end, the system sustains planar wave propagation similar to that which would be observed in a straight tube with two closed ends. The measured impulse response of the 1D model is given in Figure 4.1.

Peaks are labelled such that $f_{x,y}$ corresponds to the resonance with modal number x in length, and y in width. Clearly, y bears no meaning in a 1D model and is equal to zero. Comparison can be drawn with lengthwise modal frequencies $f_{x,y,z}$ calculated from the universal modal equation (2.29), with $y = 0$, $z = 0$ and $l = 17.6$ cm. The first five peaks are $f_{1,0,0} = 974$ Hz, $f_{2,0,0} = 1948$

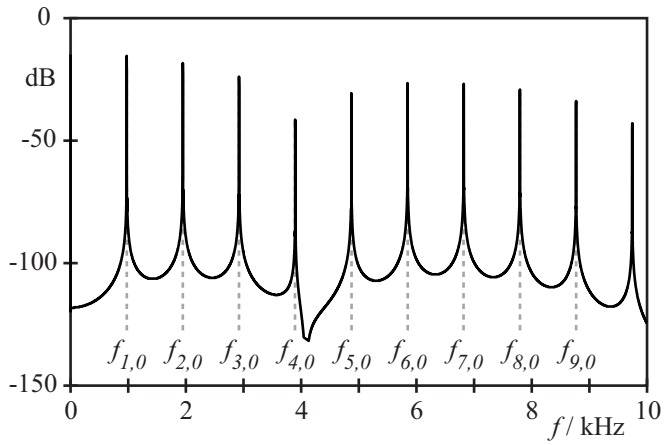


Figure 4.1: *1D straight tube impulse response*

Hz, $f_{3,0,0} = 2923$ Hz, $f_{4,0,0} = 3897$ Hz and $f_{5,0,0} = 4384$ Hz. The resonances of the 1D model in Figure 4.1 fall exactly on top of the theoretical values. This is because the 1D waveguide method provides an exact travelling wave simulation for a band-limited input signal, and is therefore an accurate model of lossless planar propagation.

4.2.2 2D Mesh

Figure 4.2 depicts a 2D rectilinear DWM model of a rectangle. The 88×20 grid of 2 mm waveguides forms a 17.6×4 cm rectangular mesh, sampled at $f_s = 242.5$ kHz. This represents a 2D plane of length 17.6 cm and width 4 cm. It consists of 1653 standard four-port junctions employing the 4-port scattering equation (2.119) and 212 one-port boundary junctions. A triangular DWM of the same size, consisting of 1914 scattering and 220 boundary junctions, has also been constructed for comparison.

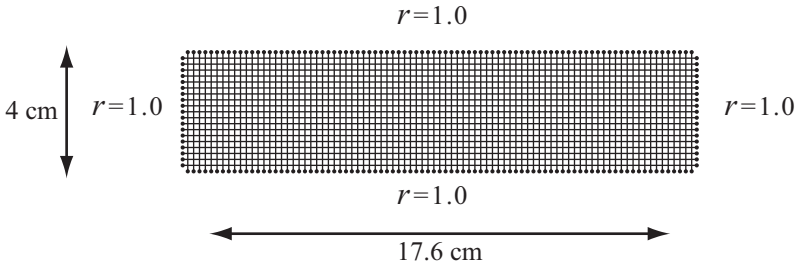


Figure 4.2: *2D Rectilinear DWM model of a rectangular plane*

In order to highlight the modal resonances, all mesh edges are configured to give maximal positive reflections $r = 1.0$. The frequency response of the 2D digital waveguide mesh model in rectilinear and triangular form is depicted in Figures 4.3 and 4.4, respectively. Results were generated using an impulse injected onto the mesh at a short distance away from one corner, with the output taken from the opposite corner. This configuration was used to ensure detection of as many resonant modes as possible.

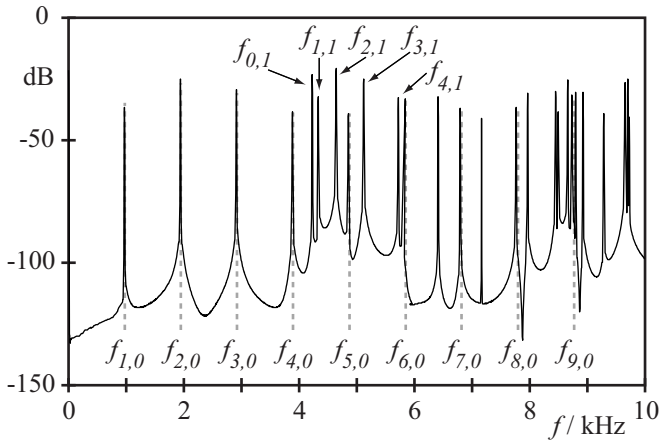


Figure 4.3: *2D rectilinear mesh impulse response*

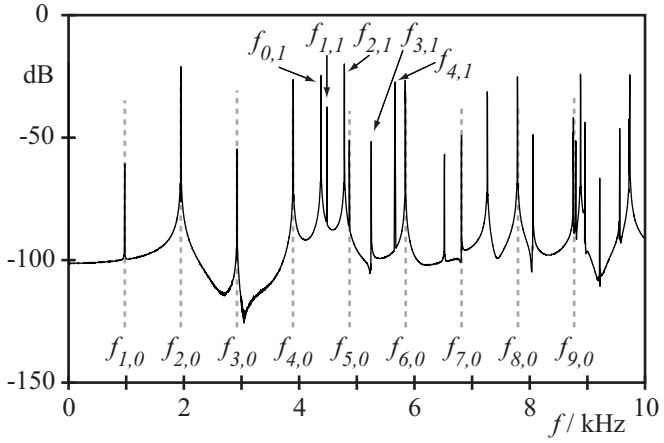


Figure 4.4: *2D triangular mesh impulse response*

The resonant peaks can be directly compared with those predicted from theory using the universal modal frequency equation (2.29). Lengthwise modes $f_{x,0}$ are highlighted by the dotted lines. In general, peaks measured from the rectilinear and triangular 2D models are in good agreement with

theory. At higher frequencies the loss of exactitude caused by the dispersion characteristics of the rectilinear mesh becomes apparent. Higher modal peaks exist at increasingly erroneous values. However, this error is small at only 1.4% away from the theoretical value for $f_{0,10}$, and 3.5% away from $f_{0,20}$ towards the top end of the audible spectrum at 20 kHz. As such they are apparent only on close inspection, and not visible in Figures 4.3 and 4.4. The high sampling frequency, and hence fine resolution of the mesh provides accurate scattering and therefore minimal dispersion effects. As discussed in Section 2.6.6, the improved dispersion characteristics of the triangular mesh increases its propagational accuracy. The maximum measured error away from theory observed in the triangular mesh simulation was found at $f_{0,20}$ to be 1.3%.

The main difference between the 1D (Figure 4.1) and 2D (Figures 4.3 and 4.4) simulations is the presence of crosswise axial and tangential modes. Using the modal frequency equation (2.29), the lowest of these should be at $f_{0,1} = 4278$ Hz, $f_{1,1} = 4396$ Hz, $f_{2,1} = 4709$ Hz, and so on. These modes can be observed from both the rectilinear and triangular mesh simulations. Each of the measured peaks were found to be accurately modelled, with less than 1% error. These are not produced by the 1D model because it only simulates the lengthwise modal pathways. It is clear that these peaks are introduced into the frequency response with the use of higher dimensionality in the model.

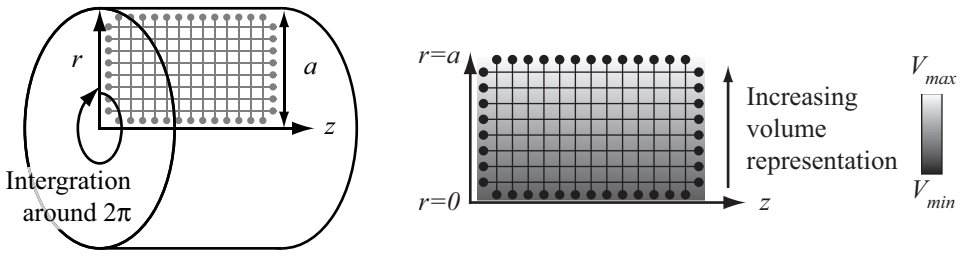
4.3 Modelling a Cylinder as a 2D plane

In constructing a 2D model of a 3D real-world system it is clear that some aspects of the problem domain will be lost. It should be decided how the 2D plane will represent the 3D tract, and which properties will be omitted. Initially, it is convenient to consider the tract as a tube, or cylinder.

4.3.1 Radial Mesh

A mesh could be constructed such that it forms a model of the plane along the cross-sectional *radius* of the tube from the centre to the inner wall. It would

constitute a 2D model of a 3D space defined in cylindrical polar coordinates. A brief outline of different coordinate systems is given in Appendix A.2. Such a model would simulate propagation along the length of the cylinder in the z-axis, and radially in the r-axis. The effects of the θ-axis could be taken into account when defining the space that is modelled by the 2D mesh, rather than simply disregarded. Figure 4.5(a) shows the manner in which a 2D DWM might be used to form a radial representation that encompasses a summation of the space defined by the angular axis θ.



(a) Radial mesh (b) Volume represented by mesh

Figure 4.5: *2D DWM cylinder model*

The summation around $0 < \theta < 2\pi$ means that waveguide elements close to $r = a$ in the mesh embody more of the cylinder volume than those closer to $r = 0$, because the of the larger circular circumference. Therefore the quantity of propagational space that is represented by the mesh increases proportionally with r, away from the centre. This notion is illustrated in Figure 4.5(b), where the radial mesh is overlayed with an volume map which increases with r towards a. Lighter grey shading represents a higher volume.

Figure 4.6 demonstrates how a radial 2D DWM vocal tract model might be constructed. An /i/ vowel 11-cylinder analogy is included to demonstrate the space represented by the mesh. This is coarsely spatially sampled from the 1D area function for visualisation purposes. Clearly, a more accurate vocal tract radial mesh model would be achieved if area functions of greater detail were used.

The manner in which the mesh would accommodate the increasing volume within the waveguides approaching the inner wall has yet not been

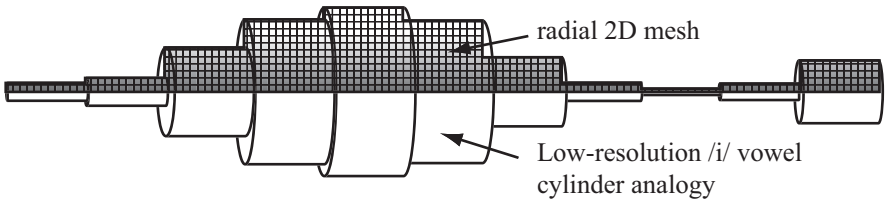


Figure 4.6: *Radial 2D DWM vocal tract model*

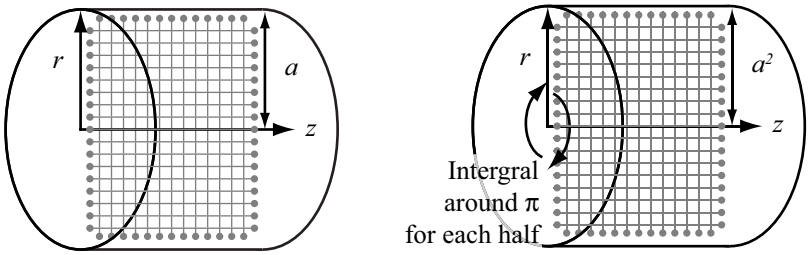
defined or tested. The additional r factor gained when defining cylindrical 3D space as a 2D mesh serves to highlight an interesting point, and a potential direction for investigation. Appendix A.2.1 shows how translation from integration in 3D Cartesian coordinates to integration in cylindrical polar coordinate systems also produces an additional r multiplication term. It is thought that adjustment of the tube radius values which are used to define the mesh size could be adapted to enhance the effects of additional volume represented in it. For example, the squaring of the tube radius to give a mesh width of a^2, rather than a, would enhance the effects of the changes in cross-sectional area along the cylinder. The difference in minimal and maximal values of a squared area function would be increased and therefore the changes that they introduce to the resonant behaviour would be accentuated. Alternatively, impedance values could be used to introduce the additional volume. A linear impedance gradient with respect to r from the the tube centre to the inner wall may also be worth investigation.

At this stage, however, no mathematical proof can be offered to justify these methods of interpreting the missing volume. However, experimentation with such factors of the model may prove to be of use in the prototype development stage, and may contribute towards an eventual formal definition.

4.3.2 Widthwise Mesh

A mesh could also be constructed that forms the 2D plane across the tube cross-section diametrically. The plane extends across the *width* from the inner wall through the centre to the opposite inner wall.

Figures 4.7(a) and 4.7(b) indicate that this configuration might be interpreted



(a) Slice through a 3D tube (b) Two half-circle volume integrals

Figure 4.7: *Diametral mesh: Two interpretations*

as a representation of a 3D tube in one of two ways. Firstly, the widthwise mesh can be thought of as a slice though the diameter of the 3D tube, which completely disregards the effects of the semi-circular tube volume either side. This viewpoint is illustrated in Figure 4.7(a). Such a model would not prove a highly accurate representation because of the missing volume. Nevertheless, it would be useful as a precursor to a 3D model, determining the potential of the technique, without implementing full 3D characteristics. In this case, the width $W(x)$ of the mesh can be said to be proportional to r, the radius of the equivalent cylinder in the 1D model.

$$\begin{aligned} W(x) &= 2\sqrt{\frac{A(x)}{\pi}} \\ &= 2r \end{aligned} \tag{4.1}$$

Figure 4.7(b) shows how the mesh across the width of the tube could be interpreted as a dual-integral mesh, similar to two of the radial meshes outlined in Section 4.3.1, connected to one another along their length. Each can be thought of as a summation of the space contained in the volume around *half* of the circular cross-sectional area. In other words, the two halves of the mesh contain a representation of the integral around $0 < \theta < \pi$, and $\pi < \theta < 2\pi$ of the circular cross-section, respectively. As with the singular radial mesh, the manner in which the additional volume could be included within the waveguides is not yet certain. Impedance gradients, or a squared radius function where the mesh width is proportional to r^2 (as suggested for

the radial mesh in Section 4.3.1) could be used to attempt to include the effects of the additional volume in the model. In the latter case, the width $W(x)$ across the y-axis of the mesh is set directly as the value of the area function at x, $A(x)$, in the following manner

$$\begin{aligned} W(x) &= A(x) \\ &= \pi r^2 \end{aligned} \tag{4.2}$$

Manipulation of the area function in this way deviates from a strictly defined physical model in that it involves experimentation without a full mathematical justification. However, such considerations can be overlooked, given the extensive complexity of the real world vocal system, and the number of other approximations, simplifications and assumptions that exist in the current stage of development of the model. Exploration of the possibilities at this stage can be viewed as initial, intuitive steps towards a working prototype, and eventually, a more formally defined model.

4.3.3 Choice for Simulations

In defining the 3D space as a 2D mesh and disregarding effects of the additional dimension it is certain that some of the resonant qualities contributed by the missing volume will be absent from the model. The widthwise 2D slice model (Figure 4.7(a)) using the *diameter* of the circular cross-section - the r-based area functions - has been chosen as a starting point for simulations. Its simplicity and ease of implementation make it an intuitive base on which to build further models. Initially, the missing volume issues will not be of concern as the investigation will be a proof-of-principle analysis into a higher dimensional model. An eventual full 3D system would not make any volume omissions.

A 2D mesh model configured in this way will not fully incorporate the crosswise resonances of the tract. It was shown in Section 2.3.5 how the Bessel function describes different modal patterns on a circular cross-section of a cylinder. Such circular waveform behaviour is not fully included in

the propagational model across the width of the 2D mesh. Moreover, the real-world cross-sectional areas measured in tract scans are not necessarily circular in shape. Such assumptions are made when quantifying the tract in a 1D area function. Nevertheless, the 2D mesh serves as a proof-of-principle for the techniques in modelling the tract in such a way. As such initial mesh widths will be calculated as proportional to radius of the corresponding circular cross-sectional area.

The lack of theory behind the inclusion of the missing volume in the approaches presented in Figures 4.5(a) and 4.7(b) make them a non-ideal starting point for investigation. However, given the similarities of the widthwise dual-integral mesh in Figure 4.7(b) to the slice equivalent in Figure 4.7(a), and the ease (simply squaring the radii) with which the slice method could be adapted to facilitate the dual-integral, it is worth attempting both. Comments made in Section 4.3.1 on integration around the circular cross-sectional area do suggest that additional r factors in the area function application to the model might have some grounding in changes of coordinate and dimension. This may offer some answers on the matter of enhancing the shape changes for the missing volume inclusion, although, as already stated, no mathematical proof can be offered at this time.

Simulations, therefore, will be made using the widthwise 2D r, and r^2 models in Figures 4.7(a) and 4.7(b), respectively. These will also be referred to as the *diameter*-based (r), and *area*-based (r^2) methods, in relation to the circular cross-section property used to determine the mesh width.

4.4 Testing the Method

The ability of the model to recreate the acoustical properties of the vocal tract in the various area function configurations can be measured in three ways. Firstly, analysis of the frequency response of the model reveals the formants that it produces. These can be compared with those observed in natural speech. In general, speech varies to a large extent from person to person. Clearly a direct comparison of the simulated formants with those

measured from an arbitrary speaker will not necessarily give an accurate description of the quality of synthesis. A comparison drawn against samples of speech taken from the source of the area functions - the X-Ray [76] or MRI [77] scan subject - at the time of acquisition would provide a more useful test. However, such data is not easily obtained from studies carried out many years previously. In the following sections average formant frequency values, as shown in Figure 3.6, taken from male speakers from a range of vowels [112] are used as an approximate guide.

Secondly, formant patterns can be contrasted with existing models and with general theory on the resonant behaviour of acoustical systems. Although, the complex and varied nature of speech makes a full theoretical justification of such a model a non-trivial task. Calculation of the modal frequencies arising from cross tract propagation in a tube of varying area function is much more complicated than finding the lengthwise modes in a straight tube. In the case of vocal tract modelling it is of interest to compare the formants obtained from a 2D model with equivalent peaks obtained from the well-established 1D model using the same area functions. In the following simulations, the 1D model used for reference is spatially sampled at 1 mm, giving sufficiently high resolution for a fair comparison.

Thirdly, the audible output generated with the application of a glottal input to the model can be compared in informal perceptual observations. This gives a measure of how similar a vowel sounds to that which would be observed in human speech. Analysis of the likeness of simulated vowels to their real-world equivalents in this work is based on the opinions of the author. It is intended to give descriptive assistance in discussions about the naturalness of synthesis offered by the 2D mesh vocal tract model.

4.5 Area Function Data

During speech, the majority of tract shape variations take place in the mid-sagittal plane. This is the side-on view, as shown in Figure 3.1, in which a large proportion of the movements of the jaw and tongue are observed.

Clearly, the use of full 3D MRI scans would be preferential in constructing a multidimensional physical model. However, due to lack of available data, reduced 1D area functions [77] were used in this project. Such data actually represents a series of cylindrical tubes connected along a straight axis, as it lacks the detail of some of the intricate cross sectional shapes present in the original scans. Despite this, the 1D data is sufficient for the purposes of this proof-of-principle system. Hence the DWM tract model discussed in this section, constructed using a 1D area function, forms a representation of wave propagation through the 2D mid-sagittal plane of the straight-axis piecewise acoustic tube analogy. Because of this, variations in area function set within the 2D model relate to an increased tract opening, rather than movement of a specific articulatory feature. However, the resulting synthesis research tool is constructed such that the techniques that are developed remain valid. They can easily be adapted for use in a model that is built around area function data that contains few simplifying assumptions.

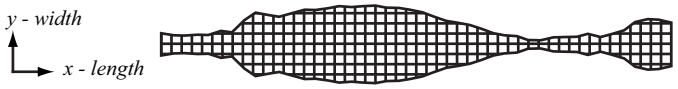
4.6 Widthwise Area Function Application

The tract shape can be applied to the width of the mesh, rather than translated into impedance as in the 1D model. The tract width is spatially sampled such that it is represented as a number of discrete waveguide lengths across the mesh. As with the 1D model, the 2D tract length determines the number of waveguides along the mesh. The following diagram illustrates the process of 2D spatial sampling, such that a waveguide mesh is constructed as a model of the 1D area function for a static vowel shape.

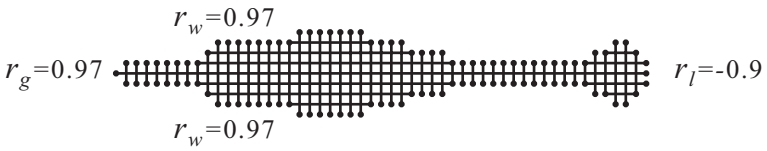
Figure 4.8(a) is the /i/ vowel area function [77]. The process of forming a 2D plane to represent the space contained within the area function is shown in Figure 4.8(b). The width W across the y-axis of the mesh is a measure of the diameter of the tube described by the 1D area function. It can also be calculated from the r^2 area function as discussed in Section 4.3.

Spatial sampling is demonstrated in Figure 4.8(c). The grid underneath the width plane is used to determine the best fit arrangement of waveguides

(a) /i/ Vowel cross-sectional area function



(b) Widthwise 2D plane analogy



(c) Spatially sampled 2D mesh

Figure 4.8: *The 2D widthwise /i/ vowel waveguide model*

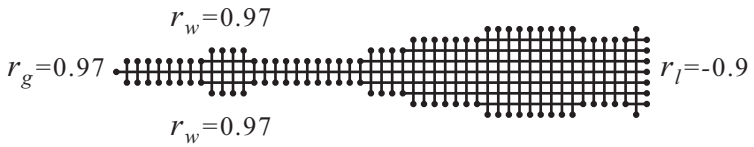
to model the acoustic cavity. Clearly the smaller the waveguide size, the higher the resolution and therefore the greater the accuracy of representation. Boundaries are configured with wall, glottis and lip reflections to be $r_w = 0.97$, $r_g = 0.97$ and $r_l = -0.9$, respectively, as outlined in Section 3.2.5.

Using this method, the mesh represents a 2D plane through the centre of the 1D tube. It will therefore sustain wave propagation and reflection in those two planes; along the tract from the glottis to the lips, and across from inner wall to inner wall. This should increase the level of representation that the model offers over the 1D equivalent. However, this method does not include waveform interaction in the angular plane around the centre of the circular cross-section. This would form the 3rd dimension if the model were to be extended. It follows, therefore, that such a model forms only a pseudo representation of the complete tract, disregarding the effects of the additional dimension in the same way that the 1D model loses widthwise detail.

Figures 4.9(a), 4.9(b) and 4.9(c) illustrate the same process for the /a/ vowel area function. Similarly, Figures 4.10(a), 4.10(b) and 4.10(c) show the construction of the widthwise /u/ vowel mesh model.

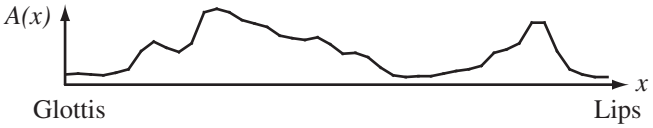
(a) /a/ Vowel cross-sectional area function

(b) Widthwise 2D plane analogy

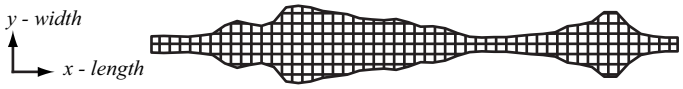


(c) Spatially sampled 2D mesh

Figure 4.9: *The 2D widthwise /a/ vowel waveguide model*



(a) /u/ Vowel cross-sectional area function



(b) Widthwise 2D plane analogy

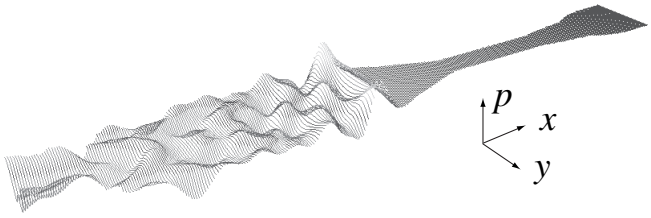
(c) Spatially sampled 2D mesh

Figure 4.10: *The 2D widthwise /u/ vowel waveguide model*

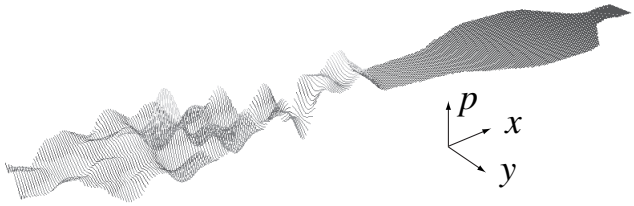
4.7 Software Implementation

The software that has been constructed to implement the DWM vocal tract model was written in the programming language C++. It forms a test-bed for the DWM vocal tract. User interaction is facilitated using the application framework MFC [113]. A Windows dialog box provides the various options for simulating the waveguide tract model in 1D and 2D, such as topology, vowel selection and reflection parameters. These parameters are set before

run-time. The model is built and the scattering equations are iterated until a *.wav* output file is saved. In this sense the software is non-realtime and non-interactive. Visualisation of the pressure waves in the tract is accomplished with the inclusion of an OpenGL [114] window within the dialog box. Figures 4.11(a) and 4.11(b) show the graphical output from the program 0.25 ms after a smoothed gaussian impulse has been applied the glottal end of the /i/ and /u/ vowel DWM models, respectively.



(a) /i/ vowel



(b) /u/ vowel

Figure 4.11: *Widthwise mapped mesh* 0.25 *ms after a smoothed gaussian impulse excitation*

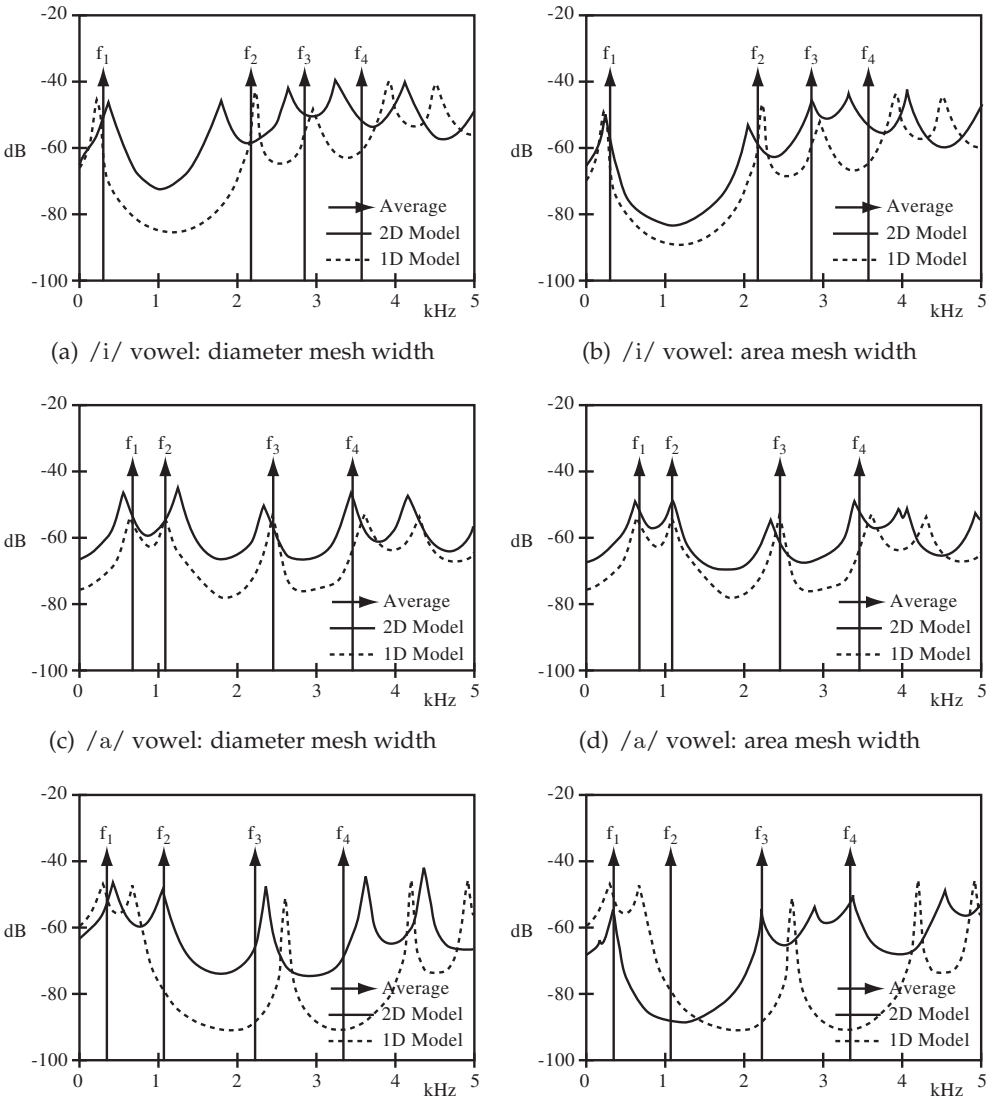
In the diagrams the input has been applied as a point source to one side of the centre of the tract in order to give a visual demonstration of wave scattering and reflection across and along the mesh. This mode of excitation results in the asymmetry seen along the tract model visualisation. It is worth noting that this is not a physically possible condition, as reflections along the tract would be symmetric along the centre line.

4.8 Simulation Results

4.8.1 Formant Analysis

The following diagrams show the formants produced by the mesh tract model using widthwise area function application. White noise is input to the mesh as a point source at the centre of the glottal end to highlight the resonant peaks in the resulting output spectrum. Figures 4.12(a), 4.12(c), and 4.12(e) show the formant pattern generated with the widthwise mapped model using the diameter based mesh width (proportional to r) (4.1) in the /i/, /a/ and /u/ vowel configurations, respectively. The graph next to each is the formant patterns generated for the same vowel, but using the area based mesh width (proportional to r^2) (4.2). Further examples are given in the second set of graphs, where Figures 4.13(a), 4.13(c), and 4.13(e) show the /ɔ/, /æ/ and /ɛ/ vowel formants generated using the diameter based mesh width. Similarly, Figures 4.13(b), 4.13(d), and 4.13(f) show the same vowel formants using the area based mesh width approach. Equivalent formant patterns generated with a 1D model with the same area functions are expressed as a dotted line. Area functions used were those obtained from MRI scans [77]. Average formant values [112] are included for reference. It is worth reiterating that, although these are not the corresponding formants to the area functions that were used, they do introduce an extra guide with which to view the formant patterns generated in the simulations.

The graphs show that the 2D model generates formant peaks that are distributed in similar patterns to the 1D equivalents. For example, the simulated /i/ vowel exhibits the low f_1, and higher and relatively close f_2 and f_3 that are seen in the 1D model. Similarly, the closely bunched f_1 and f_2 are also apparent in the /a/ vowel. The difference between the diameter and area based mesh width models can also be observed from the graphs. In Figures 4.12(a), 4.12(c), 4.13(a), 4.13(c) and 4.13(e), the 2D diameter width model gives formants at positions that are tending towards the neutral positions. That is, they are more evenly spaced than would be expected. This is because the lower order area function application has less influence in



(a) /i/ vowel: diameter mesh width

(b) /i/ vowel: area mesh width

(c) /a/ vowel: diameter mesh width

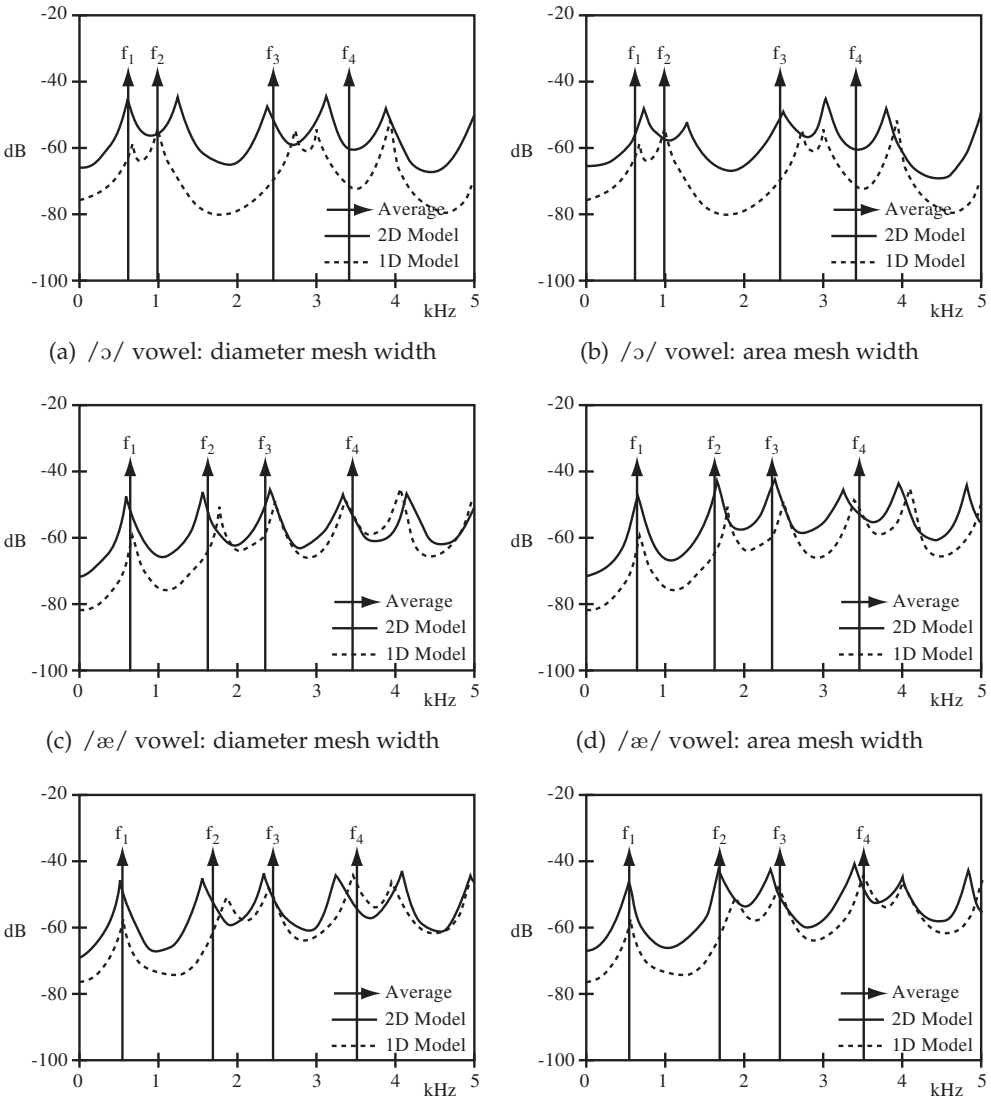
(d) /a/ vowel: area mesh width

(e) /u/ vowel: diameter mesh width

(f) /u/ vowel: area mesh width

Figure 4.12: *Widthwise mapped formant patterns - comparing diameter and area based mesh widths*

changing the resonances. Figures 4.12(b), 4.12(d), 4.13(b), 4.13(d) and 4.13(f) show that with the mesh width defined from the area, rather than diameter, has more influence in shifting the formants. The larger differences between minimum and maximum values present in the r^2-, rather than r-based, area function act to enhance the effect of the applied tract shape and push the formants closer to those generated by the 1D model.



(a) /ɔ/ vowel: diameter mesh width

(b) /ɔ/ vowel: area mesh width

(c) /æ/ vowel: diameter mesh width

(d) /æ/ vowel: area mesh width

(e) /ɛ/ vowel: diameter mesh width

(f) /ɛ/ vowel: area mesh width

Figure 4.13: *Widthwise mapped formant patterns - comparing diameter and area based mesh widths*

Clearly the 1D and 2D formant patterns are not identical. The manner in which the wave propagation is simulated in the space differs to a large extent. The peaks also do not match the average formant values. As expressed earlier, the variable nature of speech means that many different observations and measurements may arise from different speakers. Average values included on the figures are intended to demonstrate only the general extent to which the formants vary from vowel to vowel.

It seems that enhancing the effect of the constrictions for the /u/ vowel can result in an obstruction to the propagating signal. This highlights a general problem that is inherent with the widthwise mapping method of area function application. As with the other vowels, the diameter width mesh in Figure 4.12(e) shows the formants to be placed in positions that are more equally spaced than would be expected. The area based mesh width formants in Figure 4.12(f) have been overly affected by the exaggerated shape changes. The peaks have been shifted away from the neutral positions to a greater extent and it appears that the lower two formants have merged. It is considered that this can be explained with reference to the area function itself. The shape of the /u/ vowel involves some very small tract openings, such as the close proximity of the lips. If a constriction that is made to the model is enhanced in some way such that it is very small compared to the size of the resonant chambers formed elsewhere, little of the simulated signal can propagate through the restrictive opening. Formant frequencies produced by such a model might begin to show unpredictable behaviour like that seen in Figure 4.12(f).

These considerations highlight a problematic issue associated with the widthwise mapped model. If an eventual model is to simulate many of the aspects of the vocal tract, then an inability to accurately model very narrow airways will prove a strong argument against its use. Translating the area function into the distance across the mesh will always place restrictions on the minimum width allowed. In order for any propagation to take place along a narrow channel, its width must be at least two waveguides across, as illustrated in Figure 4.14. This is because a central line of waveguides, with an attached boundary junction on either side is the narrowest construction possible with a 2D mesh. Fewer waveguides would entail a 1D model. Furthermore, in dynamic speech, production of plosive sounds requires a complete stop and the release of the airflow. If a waveguide mesh model were to accommodate such obstructions smoothly, a minimum width would prove a hinderance.

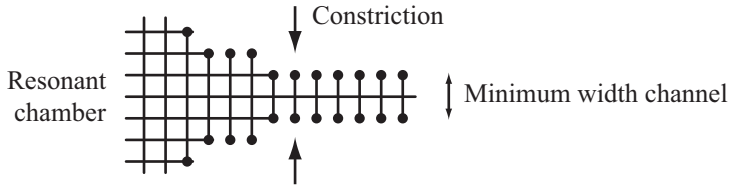


Figure 4.14: *Minimum channel defined by waveguide mesh structure*

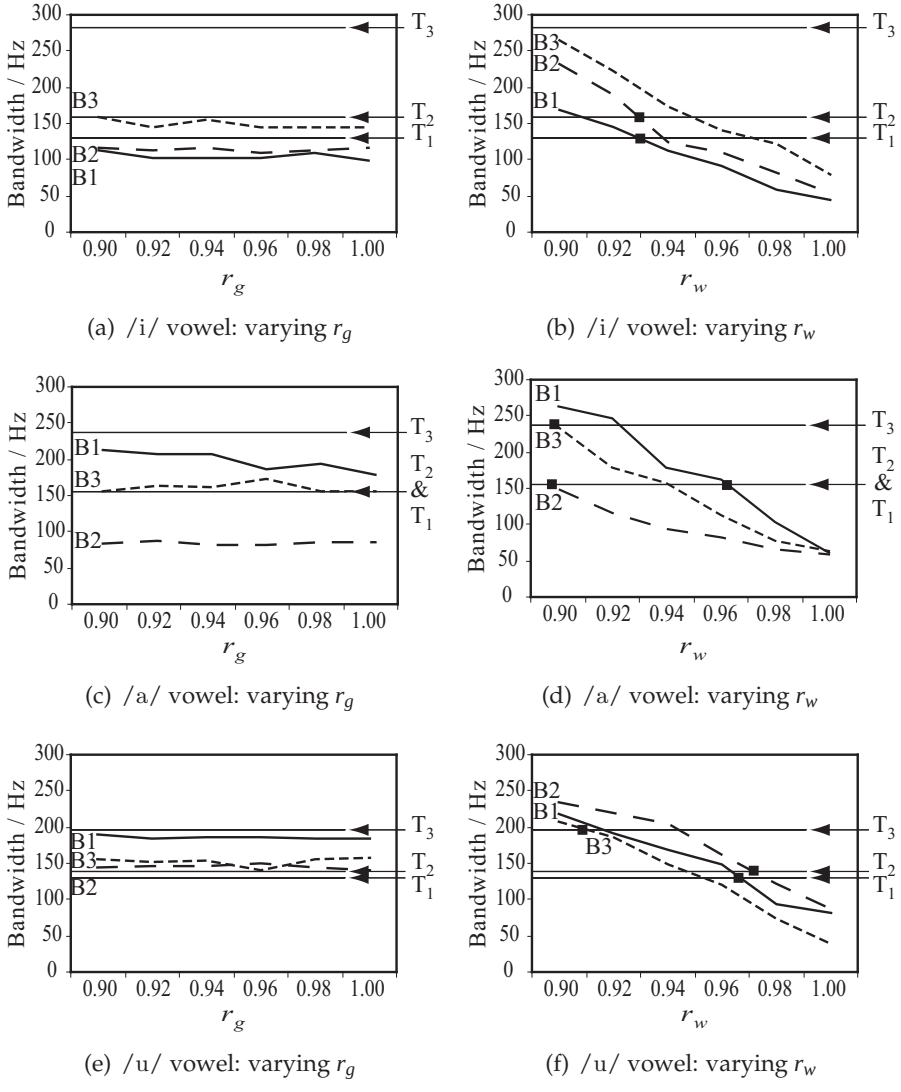
4.8.2 Formant Bandwidths

As discussed in Section 3.2.4, formant bandwidths are an important factor in speech. They define the extent to which the effect of the formant is imparted onto the excitation, and therefore directly influence the naturalness of the synthesised vowel. Large, broad bandwidths contribute little to the spectrum and result in much of the buzzing quality of the glottal pulse being audibly present in the output. Conversely, unnaturally low, narrow formant bandwidths would give rise to a metallic ringing quality in the speech output.

Formant bandwidths are dictated by internal energy reflections. In the 1D waveguide model this is governed by the boundary junction at each end. These are defined by the glottis r_g and lip r_l reflections. The 2D model extends the lip and glottis boundaries. They are modelled by multiple reflecting junctions across the mesh, the number of which depends on the opening area of the lips. In addition, the higher dimensionality introduces two extra bounding surfaces along the length of the propagation space, both characterised by the wall reflections r_w. Neglecting any frequency dependent behaviour at the boundaries, approximate reflection coefficients can be obtained from the theoretical values that were discussed in Section 3.2.5. These are $r_g = 0.97$, $r_l = -0.90$ and the wall reflections should be a highly reflective value of $r_w \approx r_g$. It is interesting to observe the influence that these parameters have over the formant bandwidths.

Figures 4.15(a), 4.15(c) and 4.15(e) demonstrate the variations observed in formant bandwidth when changing r_g and keeping $r_w = 0.97$ and $r_l = -0.9$ for the /i/, /a/ and /u/ vowel DWM models, respectively. Notation is arranged such that bandwidths $B1$, $B2$ and $B3$ correspond to the first three formants for each vowel. Target values predicted by averages taken from measurements

(see Figure 3.6 [79]) are indicated by $T1$, $T2$ and $T3$. A small black square is placed where a synthesised bandwidth matches the average, indicating a successful simulation. An alternative approach to bandwidth control is also included in the diagram, where Figures 4.15(b), 4.15(d) and 4.15(f), indicate the bandwidth response with respect to changes in r_w, and keeping $r_g = 0.97$. The results were obtained with application of random noise to a point source at the centre of the glottis end.



(a) /i/ vowel: varying r_g

(b) /i/ vowel: varying r_w

(c) /a/ vowel: varying r_g

(d) /a/ vowel: varying r_w

(e) /u/ vowel: varying r_g

(f) /u/ vowel: varying r_w

Figure 4.15: *Formant bandwidth variation in the 2D widthwise mapped mesh model*

Where bandwidths are considered with respect to r_g, none of the target values are achieved. In most cases the simulated bandwidth B is between

$50 - 100$ Hz lower than T. Some, however, such as $B1$ and $B2$ for the /i/ vowel, and $B2$ for the /u/ vowel are close at less than 25 Hz away. It can also be observed that little variation is present across the range tested. This arrangement serves as a parallel to the 1D model, where only two boundary junctions define the systems inner-reflections. Values set at each end have to accommodate commuted losses along the length. As such, small variations in the boundary coefficient have little effect in relation to the total accumulated losses that they represent. In other words, they present a low sensitivity. The 2D model includes the tract inner wall reflections separately. The glottal boundary across the mesh width is much smaller than the wall boundary along the length, and contains many fewer junctions. Bandwidth variations show low sensitivity when the smaller, uninfluential r_g is used as a control. Target values may be achieved with this configuration, although in order to increase the bandwidths to required levels greater energy loss must be arranged in the mesh. Set into the glottal and lip ends, such energy loss results in unnaturally low reflection coefficients that deviate from the theoretical conditions discussed earlier.

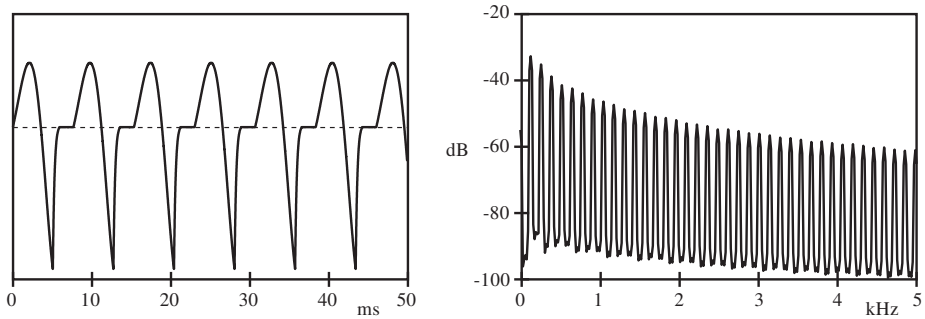
In the case where r_w is used as a controlling parameter much more variation is present. The bandwidth changes follow a more responsive, linear trend. Because of the dominating wall reflections, all but one target bandwidths are achieved within the range that was examined. The sole remaining value is $B3$ for the /i/ vowel, where the error from a successful intersection was marginal at less than 25 Hz near $r_w = 0.90$. Moreover, the target formant bandwidths were attainable whilst keeping r_g and r_l reflection coefficients set to approximate theoretical conditions. It is clear that reflection parameters set within the wall have a dominating influence over bandwidth values.

From Figures 4.15(b), 4.15(d) and 4.15(f) it can be concluded that for a reasonable match to target formant bandwidths a value of $r_w = 0.92$ should be used as an appropriate minimum-error point between success points in the 2D graph simulations. These three values conform to logical expectations in the human tract. The majority of losses should exist at the lips, where sound

is actually radiated, with some vibrational and heat conduction losses present in the fleshy inner walls of the tract, and a high reflection at the glottis.

4.8.3 Vowel Synthesis

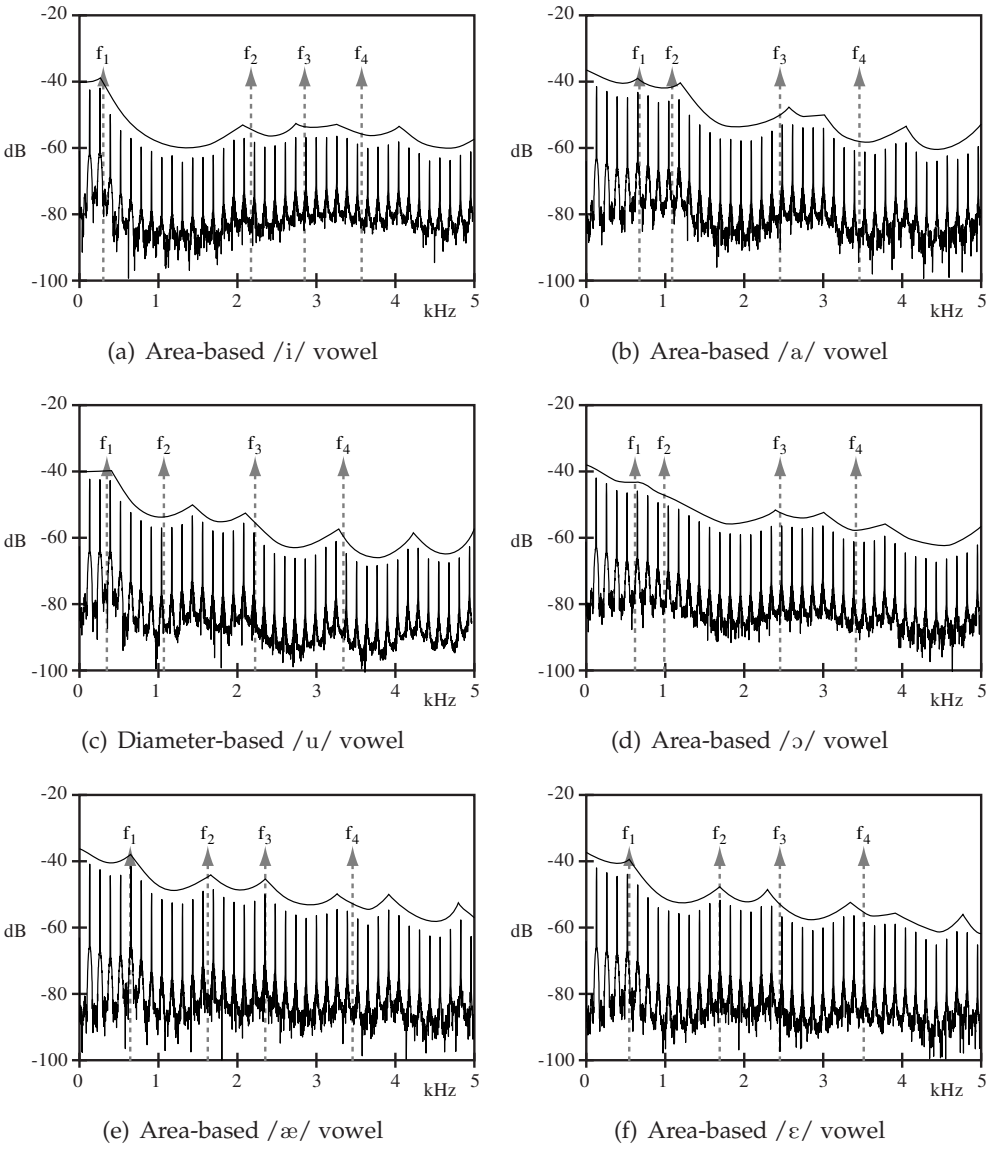
The difference between the two mesh width mapping methods is more apparent when a glottal input is used as excitation to the model in order to generate speech-like sounds. Although output generated with the diameter based (proportional to r) method sounds like the vowel that was modelled, it also bears slight audible qualities of the neutral /ə/. This is directly related to the observations made on the tend towards equally spaced formant patterns in Section 4.8.1. With the exception of the /u/ vowel, output generated with the area (proportional to r^2) based mesh width application resembles the target vowel to a much greater extent. Figures 4.17(a) - 4.17(f) show the spectra of various most natural sounding vowels generated with the application of the LF glottal waveform model as a plane source along the glottal end. The LF waveform and frequency content are shown in Figures 4.16(a) and 4.16(b), respectively. All were generated with the area-based mesh width model, apart from the /u/ vowel, which was produced with the diameter method. Approximate spectral envelopes have been included.



(a) Waveform (b) Spectrum

Figure 4.16: *LF glottal flow derivative model used for voiced excitation*

No additional natural effects, such as vibrato or pitch variation, were included as part of the glottal excitation so as to examine only the vowel resonances. The harmonics of the input signal are distinct at multiples of



(a) Area-based /i/ vowel

(b) Area-based /a/ vowel

(c) Diameter-based /u/ vowel

(d) Area-based /ɔ/ vowel

(e) Area-based /æ/ vowel

(f) Area-based /ɛ/ vowel

Figure 4.17: *2D widthwise mapped mesh vocal tract model 'best' vowel spectra*

the fundamental - 131 Hz. It can also be seen from the diagrams how the input spectrum (Figure 4.16(b)) is imparted onto the output spectrum to give it greater prominence in the lower $1-2$ kHz region. It is the opinion of the author that the vowel sounds present a strong likeness to each that was modelled. The overall quality of naturalness, however, is reduced with the use of a simplistic glottal model. Use of a better glottal input, such as a mass and spring model, would increase the naturalness of synthesis.

4.8.4 Triangular Mesh

The same model can be constructed using the triangular DWM. This should offer increase accuracy of synthesis because of its improved directional propagation, as discussed in Section 2.6.6. Figures 4.18(a) - 4.18(f) show the spectra generated with the triangular area-based width mapped tract model. Dotted lines included on each graph indicate the same formants generated by the rectilinear equivalent.

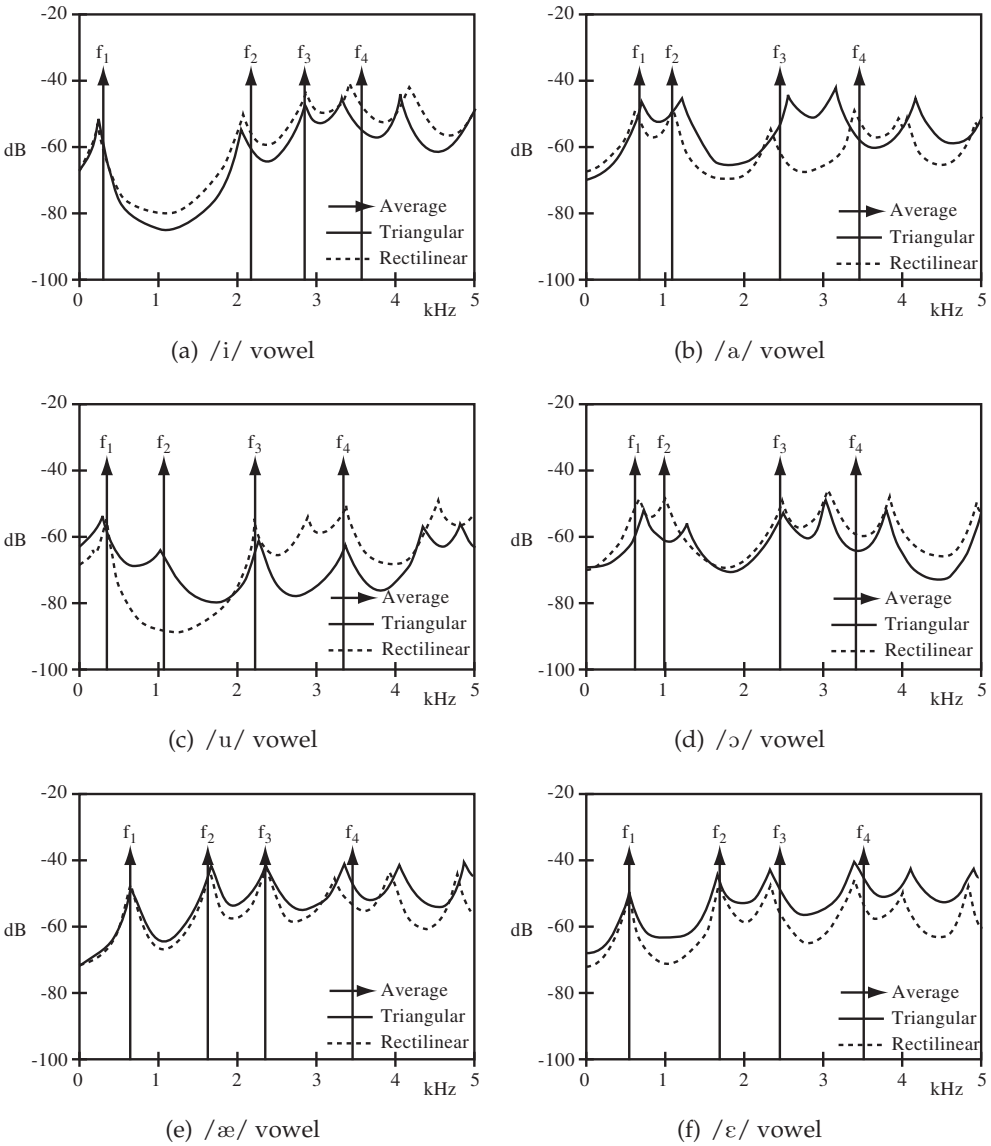


Figure 4.18: *Area based width mapped formants - comparing rectilinear & triangular mesh models*

The results from each of the two topology models correlate with expectations. Both are reasonably similar in the case of Figures 4.18(a), 4.18(d), 4.18(e) and 4.18(f). Little difference in terms of accuracy between the two topologies is visible in this low bandwidth examination. It can be seen from Figure 4.18(c) that the second formant that was absent from the rectilinear /u/ vowel mesh is present in the triangular equivalent. This is considered to be a direct consequence of using the the triangular mesh. Its construction is more capable of accommodating the boundary changes around small constrictions than the rectilinear mesh, which can only support stepped edges at 90° from one another. As such, area functions with narrow channels may be better modelled by the triangular mesh.

4.9 Dynamic Behaviour

The vocal tract shape constantly varies during connected speech. Movements of the tongue, lips and jaw amount to changes in the area function that alter the resonant properties of the tract and create the different speech sounds. An ability to recreate this dynamic behaviour is essential in an articulatory synthesis model.

On a simplistic level, a diphthong can be modelled as a linear interpolation between two vowel area functions. Figure 4.19 shows the how the formants of the 1D waveguide tract model change with a slide from the /i/ to /a/ vowels. The transition is applied smoothly over 500 ms. The change in formants is clear.

The 2D rectilinear mesh can also be used to facilitate a vowel slide. A linear interpolation between the /i/ and /a/ vowel diameter based mesh-width vocal tract models is shown in Figure 4.20. The overlayed dotted line shows the 1D equivalent from Figure 4.19.

Although it produces accurate formant synthesis, the widthwise mapping approach to area function application does not fully accommodate dynamic changes in tract shape. The vocal tract configuration used to generate the /i/ to /a/ vowel slide in Figure 4.20 involves a widening in mesh width at the

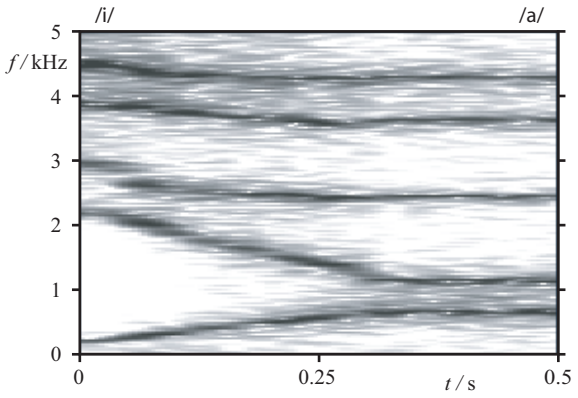


Figure 4.19: *1D tract model /i/ to /a/ vowel slide*

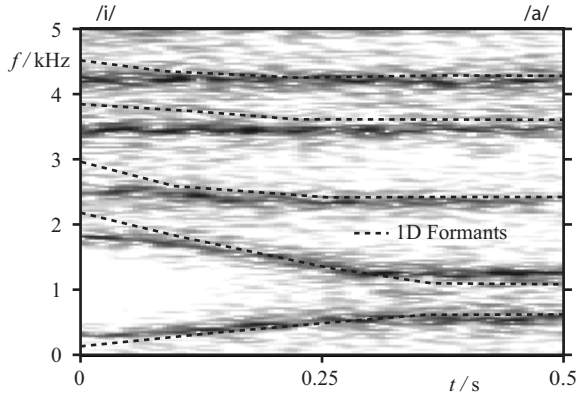


Figure 4.20: *2D mesh width tract model /i/ to /a/ vowel slide*

mouth, and a narrowing towards the middle. Referring to the mesh layout around these areas in Figures 4.8(c) and 4.9(c) it is clear that this transition will require additional waveguides to be added around the mouth, and removed from the middle region. These changes force surrounding junctions to alter their behaviour. A small movement of a mesh boundary is examined in Figure 4.21. The two states of the mesh highlight the difference in boundary structure arising from a small change in modelled width. Moving from the left to the right state illustrates a narrowing of the tract, such that the grey junctions and waveguides are no longer active. Similarly, moving from right to left demonstrates the reverse process where the grey junctions and waveguides are added to the mesh.

This dynamic restructuring of waveguides at run-time can be problematic in maintaining the continuity laws governing the mesh scattering equations.

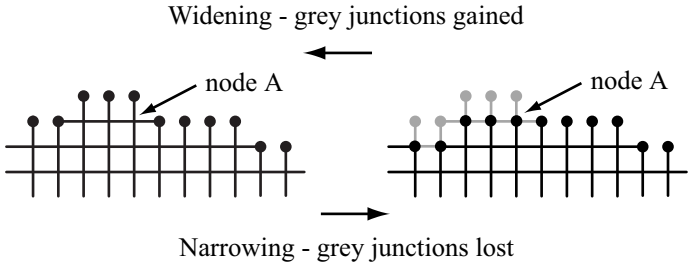


Figure 4.21: *Junctions gained or lost in a mesh boundary movement*

For example, an increase in width might see a 1-connection boundary junction, such as node A on the right hand side of Figure 4.21, suddenly take on the role of a 4-port scattering junction. This would lead to the averaging of the single incoming pressure value across four outputs and hence a sharp step in pressure gradient in the new mesh area.

Some attempts were made to develop a junction with the capability to accommodate such changes. Scattering equation (2.94) was reconfigured such that errors introduced by additional or removed pressure inputs were spread evenly across existing connections as to minimize their effects. It was found, however, that the manipulation of the equations was often in contradiction with the underlying acoustic theory, and hence introduced more instabilities rather than fewer. Minimum disruption to the pressure balance at each junction was achieved simply by defining new pressure components to be set to zero and that lost pressure components are disregarded.

The distance involved in moving a boundary between minimal and maximal area function values is about 20 mm. The changes are infrequent, as the number of junction manipulations is negligible when compared to the number of samples in the given duration for the transition. For example, using (2.118) the waveguide size in a high resolution mesh sampled at 120 kHz is about 4 mm. In the approximately 500 ms required for the transition in a diphthong, this would result in five junction changes at each moving boundary point over 60000 samples. However, the small discontinuities propagate across the mesh and are still audible in the output waveform. Figure 4.22 shows the output generated for 30 ms after a boundary change is initiated in the 2D model during a vowel slide. The LF glottal flow derivative

model was used as excitation [89]. Approximately four cycles of the output waveform are shown after a boundary alteration instant. Discontinuities are clearly visible about 14 ms after a step in boundary movement, which are audible as a high frequency click in the output.

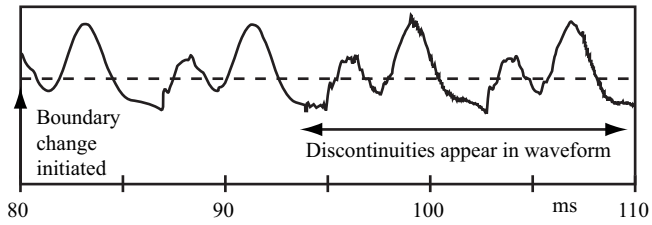


Figure 4.22: *2D mesh width dynamic changes: discontinuities in the waveform*

4.10 Conclusions

This chapter has presented and analysed a DWM model of the vocal tract. A 2D model is implemented that builds upon the well established 1D waveguide chain method. Two methods of defining the mesh width were suggested; a diameter-based method which uses a width that is proportional to the radius of the tract, and an area-based scheme which is proportional to r^2. The mesh used for simulation of the space contained within the tract includes transverse and longitudinal propagation and therefore incorporates more of the resonant characteristics. It was found that accurate formant synthesis may be achieved. In general, the formant patterns are observed as close to those generated with the 1D model. The simulated vowels are audibly similar to those observed in natural speech, with those generated with the area-based width giving increased likeness. The perceived naturalness of the vowels would be further enhanced with the use of a better glottal model.

Analysis of the formant bandwidths has shown that the additional boundary along the wall of the mesh provides a useful tool in regulating energy losses. Using r_w as a controlling parameter, a linear bandwidth response is obtained. This facilitates fine-tuning of formant strength.

Lastly, the dynamic ability of the DWM model was discussed. It was indicated that the spatially sampled nature of the static waveguide structure

is not ideal for the modelling of dynamic systems. The discontinuities that are introduced into the waveform when moving the boundaries of the mesh tract model were demonstrated. This lack of stable dynamic capability is an important issue that will greatly limit the use of the 2D DWM as a vocal model, or indeed as a model of any moving acoustic system. The following chapter will look at novel ways in which to address this problem.

Chapter 5

A Dynamic Real-Time Approach

5.1 Introduction

The spatially sampled nature of the digital waveguide mesh means that it models a static structure. Problems associated with dynamic manipulation of the DWM vocal tract model were highlighted in Section 4.9. Audible waveform discontinuities were introduced into the resulting output. This chapter details a novel approach for introducing the shape changes contained in the area function to a 2D mesh model of the vocal tract. A rectangular mesh is used to represent a straight tube. Changes to the tract shape are implemented using waveguide impedance values, rather than with the unstable method of moving the mesh boundaries. The mesh remains rectangular and is therefore not prone to discontinuities arising from junctions that are forced to modify their scattering properties. It is worth noting that the model is discussed here in impedance terms in Z, whereas waveguide scattering equations were derived in Section 2.5 using admittances in Y. The two are interchangeable, given the reciprocal relationship that exists between them $Z = 1/Y$.

To begin with, the notion of manipulating the resonances of a straight tube mesh with impedance changes is introduced. Two methods of applying a constriction across the width of the mesh are described. Linear and raised-cosine functions are presented for impedance increases towards the edges of the mesh to bring about the constriction. Validation of the technique is then

given in the form of the changing modal frequencies of a straight tube when a steep constriction is slowly applied to the mid point of a mesh. The manner in which the method can be used to create impedance maps based on 1D area functions is then discussed. Next, a brief introduction to the software that has been developed to test the models is given. Simulation results are given in the form of a formant frequency analysis, followed by spectra of some synthesised vowels. Lastly, the dynamic and real-time ability of the impedance mapped mesh is considered.

5.2 Impedance-Based Area Function Application

An alternative method of imparting the effects of the shape changes contained in the area function onto the 2D mesh has been developed. In the well-established 1D model the area function is translated into waveguide impedance. Building on this technique, it was considered how such changes could be made to the impedances within a static rectangular 2D DWM, so as to have the same airflow-constriction effects along the tract.

A step in impedance in the 1D model gives rise to some amount of transmission through the discontinuity and some amount of reflection back in the incident direction. Similarly, a step in waveguide impedance across the width of a rectangular 2D mesh, such that the change is experienced by a signal travelling along the length, would produce forwards-backwards reflections. However, with equal impedance across the mesh at each impedance step, the area function change would have little effect in the cross-wise plane. This would render the use of the increased dimensional representation somewhat superfluous. The impedance discontinuity can be applied such that it is not uniform across the mesh. A constriction can be therefore be applied as raised impedance regions if allowances are made in order to encourage cross tract reflection against them. A lower central channel is defined such as to act as a direct lengthwise propagation pathway. Constrictions are applied as increases in impedance at the edges of the mesh, so as to facilitate cross-reflections. Figure 5.1 illustrates the process of introducing raised impedance

regions towards the edges of the straight tube mesh as to alter its resonant behaviour.

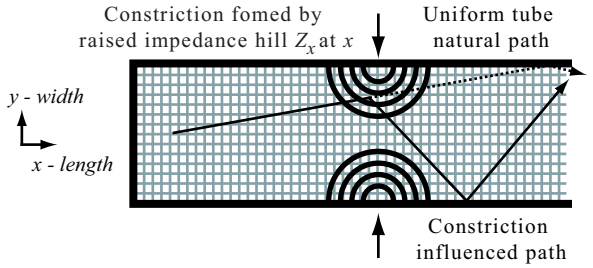


Figure 5.1: *Raised impedance hills causing a constriction*

The semi-circular contours on the mesh represent a transition from the lower, default impedance value to the higher regions towards the edges at the point of constriction. With a given area function $A(x)$, Z_{min} is defined as the minimum impedance arising form the maximum area value.

5.2.1 Translating Tract Radius into Waveguide Impedance

In the 1D model, the impedance value Z_x at each spatial sampling instant is calculated from the 1D area function $A(x)$ using (2.11). The mapping of the area function onto waveguide impedances in the 2D DWM tract model is a new technique which has not yet been formally defined. There is, therefore, scope for experimentation with the model and the manner in which it is used to represent the tract space. The impedance transition across a constriction such as that illustrated in Figure 5.1, will be between Z_{min} at the centre of the mesh, leading to Z_x at either side. This impedance Z_x defines the size of the constriction function at each point across the mesh. How it is defined in relation to the 1D area function needs to be addressed.

r^2 Area Function

As a starting point, Z_x can be obtained in the same manner as the 1D model, directly from (2.11). It can therefore be described as being inversely proportional to the circular cross-sectional radius squared, r^2, because ρ and c

are constant in each corresponding 1D tube section. Equation (2.11) becomes

$$\begin{aligned} Z_x &= \frac{\rho c}{A(x)} \\ &= \frac{\rho c}{\pi r^2(x)} \end{aligned} \tag{5.1}$$

r^3 Area Function

The selection process for the mesh width mapping method in Section 4.3 introduced the concept of the missing volume when modelling a 3D space with a 2D plane. Linear impedance changes and additional r factors were suggested as ways of incorporating some of the effects of the omitted space into the DWM. No mathematical proof or formal justification for these suggestions has been offered. However, the augmented r-factor mesh gave improved vowel likeness in formant and vowel simulations in Section 4.8. For this reason it is worth considering how to increase the effects of the area function in the 2D impedance mapped DWM. The radius value within the area function can be raised to an additional power of itself in order to facilitate this. In this sense the impedance function can be said to follow a relationship that is inversely proportional to r^3. In this case, equation (2.11) becomes

$$\begin{aligned} Z_x &= \frac{\rho c}{A'(x)} \\ &= \frac{\rho c}{\pi r^3(x)} \end{aligned} \tag{5.2}$$

Where the tube cross-sectional area $A'(x) = \pi r^3(x)$ is not strictly a real quantity, but a manipulated form of the area function. This leads to a 2D system that isn't a true representation of the tract space. But, given the number of other approximations which must be made in forming such a model, and the lack of a rigorous mathematical derivation, it follows that deviation from strict modelling methodology can form a valid part of the investigation.

In summary, impedance maps derived from both r^2 and r^3 area functions will be used in the following tract models.

5.2.2 Constriction Function

Two functions have been considered in an attempt to find the most appropriate way to facilitate the transition. Figure 5.2 demonstrates how a linear increase can be used to define the impedance variation $Z(x,y)$ across the y-axis of the mesh at a point x along the length.

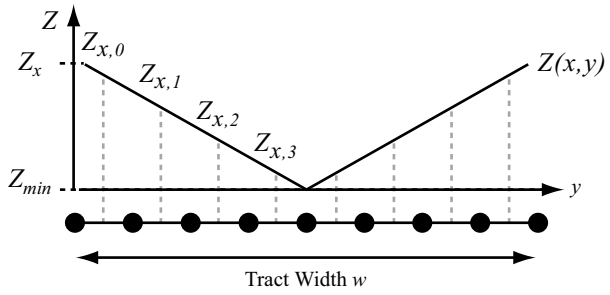


Figure 5.2: *Linear impedance hills either side of a constriction*

An impedance map configured using a linear increase towards the mesh edges is defined by

$$\begin{aligned} Z(x,y) &= Z_x - \frac{(Z_x - Z_{min})y}{w/2} && 0 \leq y \leq w/2 \\ Z(x,y) &= Z_{min} + \frac{(Z_x - Z_{min})(y - w/2)}{w/2} && w/2 \leq y \leq w \end{aligned} \tag{5.3}$$

Similarly, Figure 5.3 shows how this can also be achieved with the use of an inverted raised cosine function.

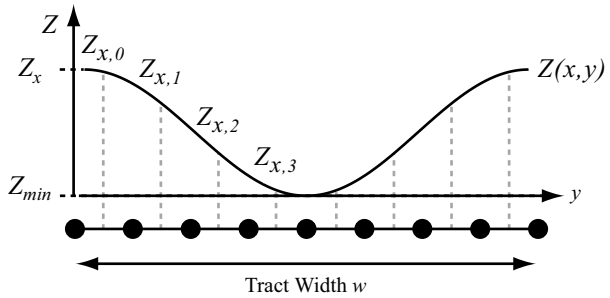


Figure 5.3: *Raised cosine impedance hills either side of a constriction*

The following equation is used to define the raised cosine impedance map.

$$Z(x,y) = Z_x - \frac{(Z_x - Z_{min})}{2}\left[1 + cos\left(2\pi(\frac{y}{w} - \frac{1}{2})\right)\right] \tag{5.4}$$

It should be noted that the two functions offered here are by no means the definitive answer in translating the area function into the impedance map. Many other functions may prove useful in further development of the method, but at this stage the linear- and cosine-based approaches are intended to demonstrate the potential of the impedance technique in dynamic 2D mesh simulations.

These methods of area function application to the waveguide mesh model also allow for simulation of a complete stop to the airflow, such as that found in plosive articulation. No minimum width channel is defined and so the constrictions that are applied to the mesh can be increased to the point that they form an obstruction to the signal propagation. Figure 5.4 illustrates a raised cosine function impedance constriction that has been enlarged to form the high impedance Z_{stop}. This relates to the impedances in Figure 5.3 in the sense that $Z_{stop} >> Z_x > Z_{min}$. The middle of the function has therefore been omitted from the graph in order to scale the diagram to allow reference with the lower impedance Z_x.

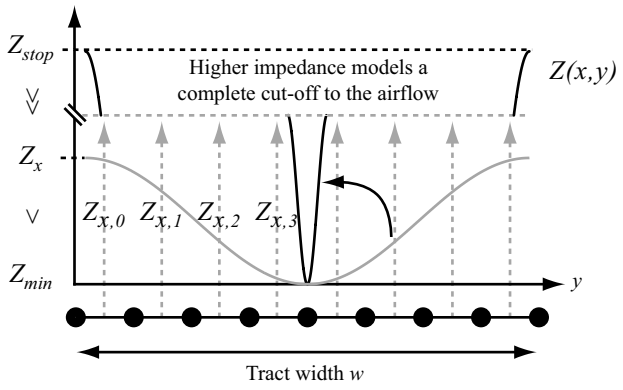


Figure 5.4: *Raised cosine function for high impedance obstruction Z_{stop}*

5.2.3 Validation

The method can be analysed with examination of the change in modal frequencies brought about by a simple change to the shape of the modelled space. Consider the 17.6×4 cm rectangular mesh of equal impedance waveguides in Figure 4.2. All four boundaries are set to fully positive

reflections to allow for comparison with modal frequencies calculated with the modal frequency equation (2.29). Half-way along, at $x = 8.8$ cm, a narrow raised cosine impedance constriction is gradually applied across the width, such that the resulting map is as portrayed in Figure 5.5.

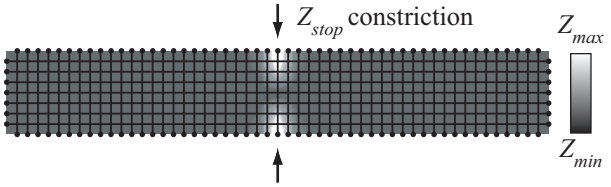


Figure 5.5: *Impedance map of a sharp constriction halfway along a straight rectangular mesh*

The high impedance constriction is applied as a linear interpolation between Z_{min} and Z_{stop}, where $Z_{stop} = 1000Z_{min}$. Excitation is of the form of a smoothed gaussian impulse at a point in the centre of the glottal end. Figure 5.6 demonstrates the change in frequency response of the mesh over the 500 ms taken for the transition.

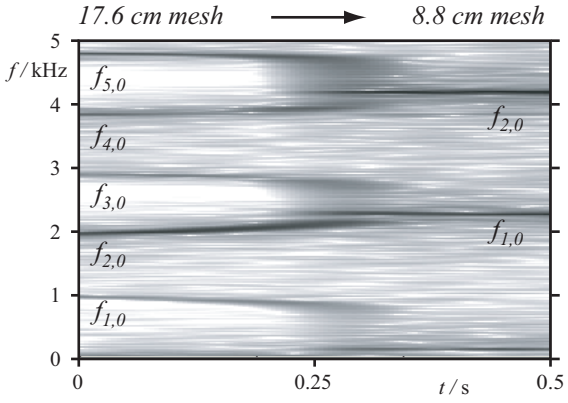


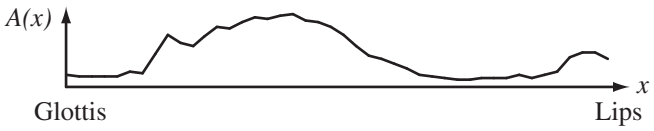
Figure 5.6: *Changes in modal frequencies of a rectangular mesh resulting from a constriction*

At time $t = 0$, the first four lengthwise modal resonances of the 17.6 cm mesh are approximately as would be calculated using (2.29) as $f_{1,0} = 974$ Hz, $f_{2,0} = 1948$ Hz, $f_{3,0} = 2923$ Hz and $f_{4,0} = 3897$ Hz. As the constriction is applied to the middle of the mesh it has the gradual effect of halving the length. To begin with, modal resonances $f_{1,0}$, $f_{3,0}$ and $f_{5,0}$ in Figure 5.6 are seen to decrease in frequency and increase in bandwidth. Peaks $f_{2,0}$ and $f_{4,0}$ experience a growth in both frequency and bandwidth. At about $t = 0.25$ ms,

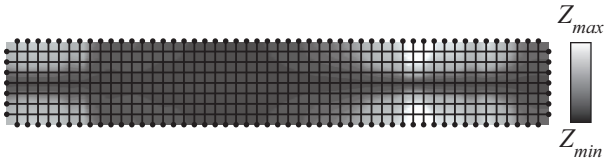
the constriction becomes an obstruction. At this point, neighbouring peaks merge. The remaining lengthwise modes are those that would be expected from a mesh of 8.8 cm. Using (2.29), these should be at $f_{1,0} = 1948$ Hz and $f_{2,0} = 3897$ Hz. Peaks in Figure 5.6 at $t = 500$ ms are observed at slightly higher values of $f_{1,0} = 2249$ and $f_{2,0} = 4204$. This can be accounted for by the fact that length of the 8.8 cm mesh is reduced to a small extent by the width of the constriction itself.

5.3 Vowel Impedance Maps

Equations (5.3) and (5.4) can be used to form a 2D impedance plane that is derived from the 1D area function. Waveguide impedances within the rectilinear mesh structure are set according to this map. Figures 5.7(a), 5.8(a) and 5.9(a) give 1D area functions taken from MRI scans [77] of the vocal tract in the position held for the /i/, /a/ and /u/ vowels, respectively. Beneath each of these is the corresponding 2D r^2 impedance map generated using the raised-cosine impedance function (5.4). Lower impedance regions on the map are indicated by darker shading, whereas higher impedances are represented by the lighter areas. Equivalent linear impedance maps are not included for reference as little visible difference is apparent at the scale chosen for graphical representation.



(a) /i/ Vowel cross-sectional area function

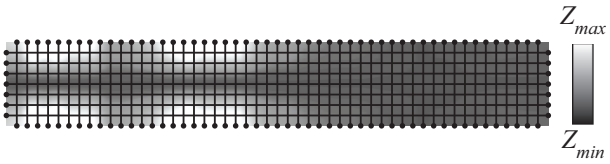


(b) Rectangular mesh overlayed with impedance map

Figure 5.7: *The 2D impedance mapped /i/ vowel waveguide model*

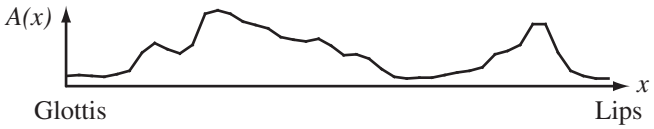
The small airway created by raising the back of the tongue in production

(a) /a/ Vowel cross-sectional area function

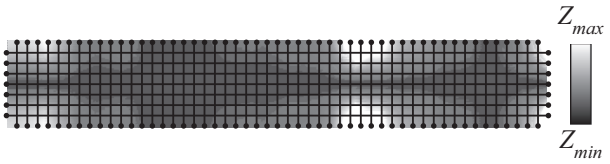


(b) Rectangular mesh overlayed with impedance map

Figure 5.8: *The 2D impedance mapped /a/ vowel waveguide model*



(a) /u/ Vowel cross-sectional area function



(b) Rectangular mesh overlayed with impedance map

Figure 5.9: *The 2D impedance contour /u/ vowel waveguide model*

of the /i/ vowel is apparent as the narrow darker impedance channel surrounded by large lighter regions in the mouth area of Figure 5.7(b). Similarly, the open mouth position held for production of the /a/ vowel can be seen as the large, darker, lower impedance region in Figure 5.8(b).

5.4 Software Implementation

Software has been developed to demonstrate the capabilities of the impedance mapped rectilinear DWM vocal tract model. A dialog box application has been written in the C++ programming language with MFC. The open source audio I/O library PortAudio has been used to facilitate the real-time sound output form the model. A screen-shot of the application is shown in Figure 5.10.

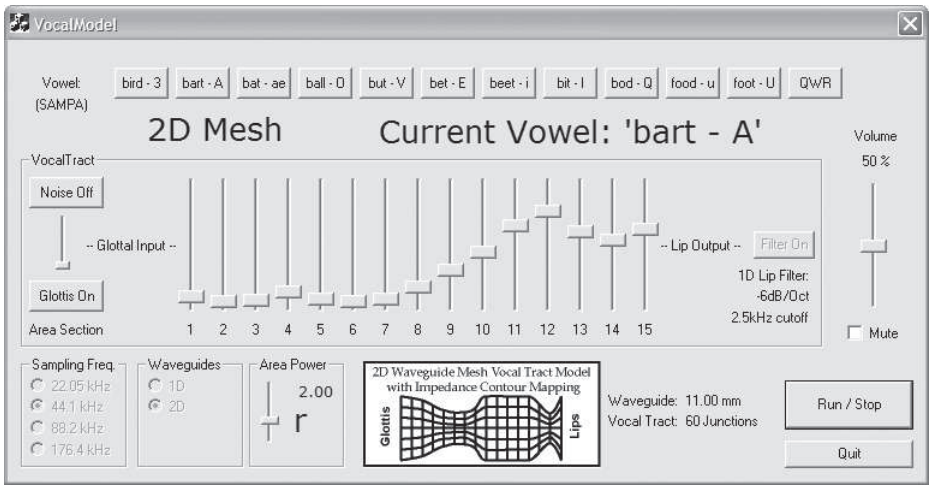


Figure 5.10: *Real Time Application*

Input is arranged in the form of an LF glottal waveform for voiced phonation, or random noise for whispered excitation. Vowel selection buttons along the top of the dialog box can be used to initiate a slide from the current vowel. The sliders across the middle allow for real-time control of the area function that is applied to the rectangular mesh. A sharp downward movement in slider position at the lip end produces a sudden decrease in area function. This can be used to simulate plosive articulation, as would be heard in the consonant /p/ or /b/.

5.5 Simulation Results

The model constructed for the following simulations was formed using a 17.6×4 cm rectilinear DWM, with a sampling frequency of $f_s = 242.5$ kHz and waveguide size of 2 mm. Boundary reflections were set to theoretical values $r_w = 0.97$, $r_g = 0.97$, and $r_l = -0.9$, as discussed in Section 3.2.5.

5.5.1 Formant Analysis

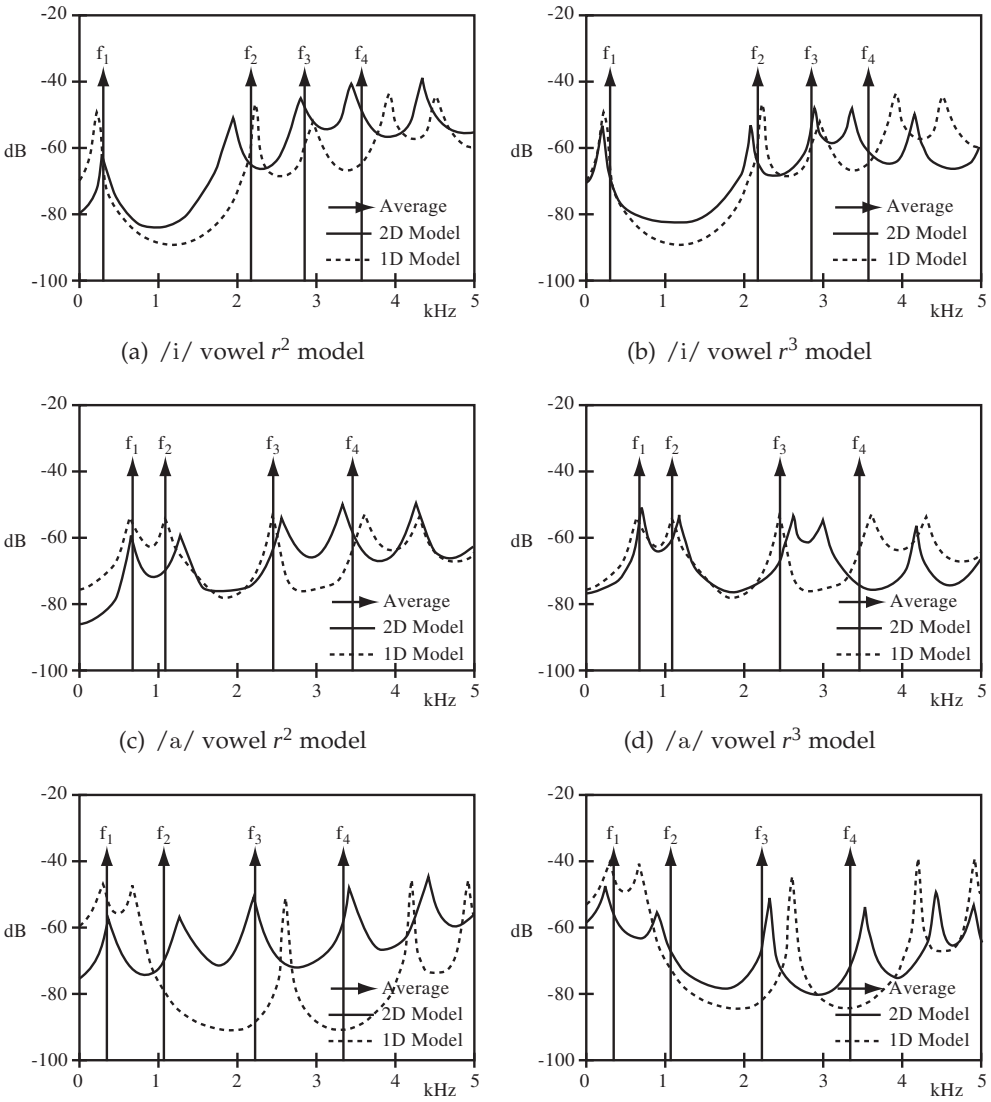
Analysis of the formant patterns generated using the raised-cosine impedance mapped method of area function application indicates its vowel synthesis potential. Figures 5.11(a), 5.11(c) and 5.11(e) show the spectral responses of the /i/, /a/ and /u/ vowel r^2 model, respectively. Neighbouring graphs

5.11(b), 5.11(d) and 5.11(f) give the enhanced r^3 model spectra for the same vowels. Excitation was applied to the tract in the form of random noise as a point source in the centre of the glottal end. The dotted lines provide comparison against spectra generated using the 1D waveguide chain model, using the same area functions. Measured formant values are also compared with the average values shown in Figure 3.6 [79]. Once again, it is worth stating that these average formant values cannot be directly compared to the measured ones, but serve as an additional indication as to the general extent that the formants change in natural speech.

Similarly, Figures 5.12(a), 5.12(c) and 5.12(e) show the formant patterns for the /ɔ/, /æ/ and /ɛ/ r^2 vowel models, respectively. Neighbouring graphs 5.12(b), 5.12(d) and 5.12(f) give the enhanced r^3 spectra for the same vowels.

As with the widthwise model in Section 4.8.1 the impedance mapped peaks do not exactly match those generated using the 1D model. The underlying signal propagation mechanism is different. An identical formant pattern would therefore not be expected. Towards the higher end of the spectrum above 4 kHz, there is less concurrence between the 1D and 2D models. However, in general a good agreement exists between the lower three or four 2D peaks and their 1D equivalents in that they are shifted in the correct direction away from the neutral positions. For example, a lowering of f_1 and raising of higher formants can be observed for the /i/ vowel. Similarly, f_1 and f_2 have been moved closer together to tend towards the /a/ vowel.

Changes in the frequency response due to enhancements made to the applied area function are clear from the difference between the r^2 vowels on the left and r^3 on the right. In some cases, such as in Figures 5.11(a) and 5.11(e) the r^2 peaks have not been shifted as far as those from the 1D models. Figures 5.11(b) and 5.11(f) show how the effects of increasing the strength of the constrictions moves some of the formants to be more in line with 1D equivalents. The other spectra in Figures 5.11(c), 5.12(a), 5.12(c) and 5.12(e) show that the r^2 vowels give a good match to the 1D model.



(a) /i/ vowel r^2 model (b) /i/ vowel r^3 model

(c) /a/ vowel r^2 model (d) /a/ vowel r^3 model

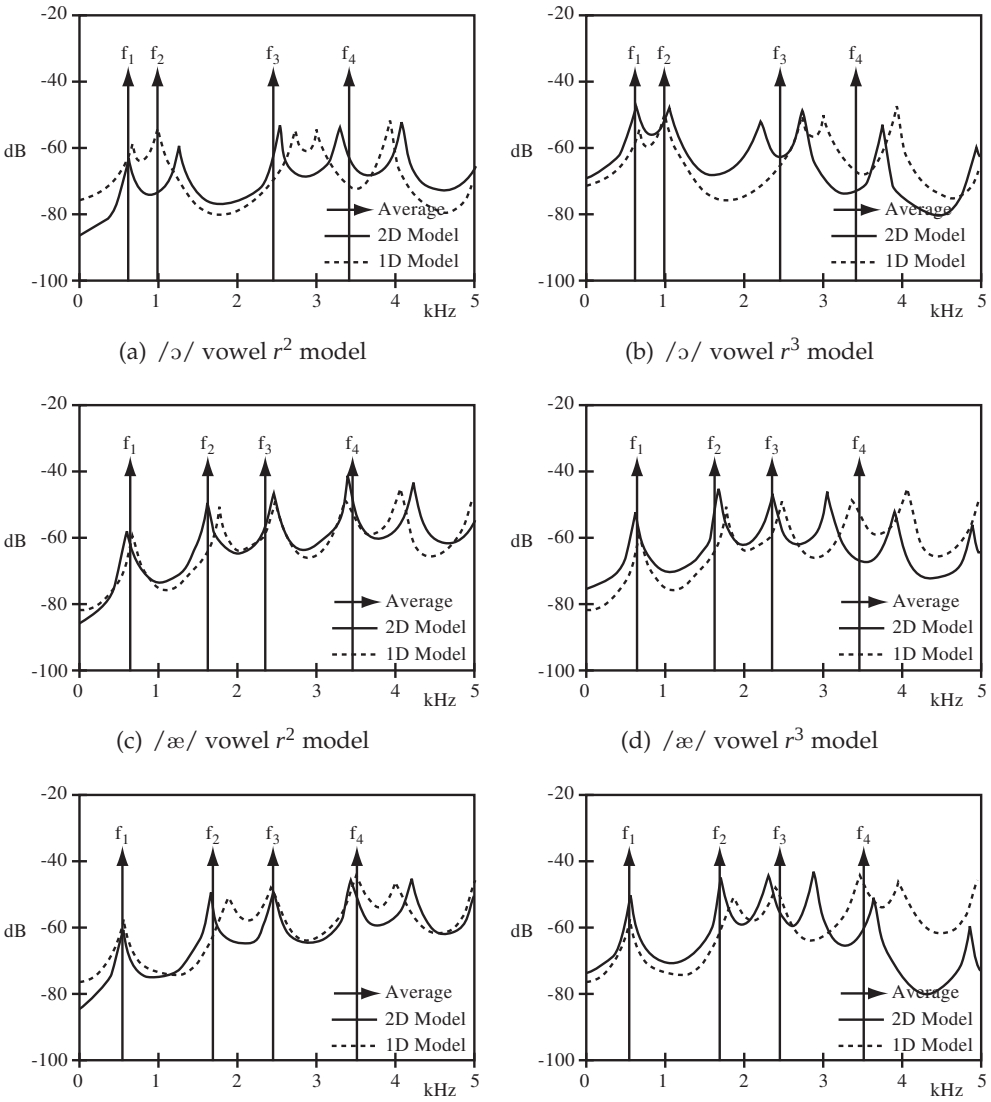
(e) /u/ vowel r^2 model (f) /u/ vowel r^3 model

Figure 5.11: *Raised-cosine impedance mapped formant patterns*

The corresponding r^3 formants in Figures 5.11(d), 5.12(b), 5.12(d) and 5.12(f) appear to have been shifted beyond the target values.

5.5.2 Vowel Synthesis

With excitation of the impedance mapped model using the LF glottal waveform (shown in Figure 4.16(a)) as a plane source along the glottal end of the tract, the following vowel spectra are produced.



(a) /ɔ/ vowel r^2 model (b) /ɔ/ vowel r^3 model

(c) /æ/ vowel r^2 model (d) /æ/ vowel r^3 model

(e) /ɛ/ vowel r^2 model (f) /ɛ/ vowel r^3 model

Figure 5.12: *Impedance mapped formant patterns*

In some cases, standard r^2 formant values were measured to be closer to the 1D model over enhanced r^3 equivalents in the previous section. However, perceptually, vowels simulated using the enhanced r^3 area function are considered to produce sounds that bear more audible similarities with the real-world equivalents. Spectra given in Figures 5.13(a) to 5.13(f) are from the r^3 model, and are considered to represent the most natural vowel synthesis offered by the impedance mapped method.

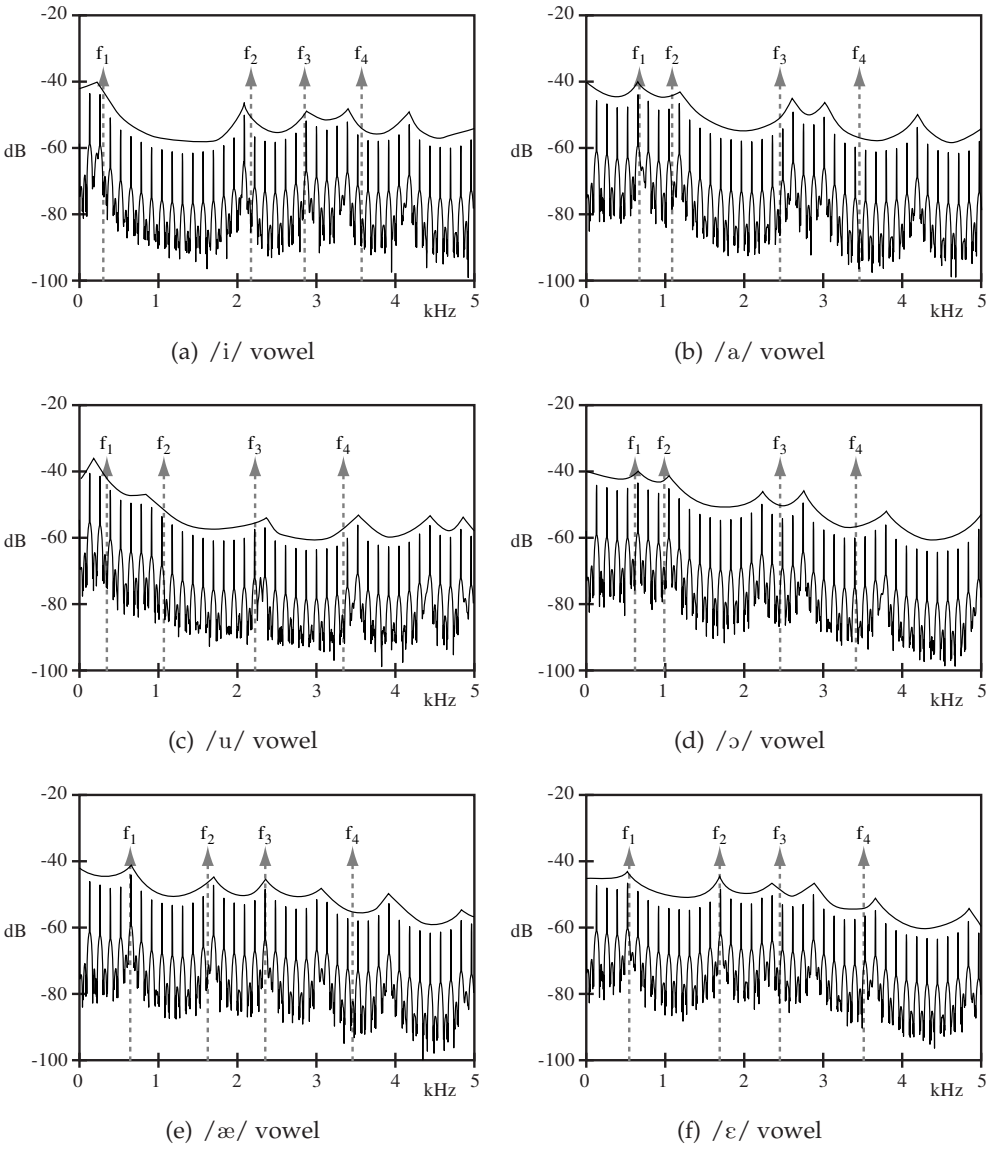


Figure 5.13: *2D r^3 impedance mapped mesh vocal tract model vowel spectra*

5.6 Dynamic Behaviour

Impedance mapping area function application allows for stable dynamic changes to be made to the mesh. This can be demonstrated with simulation of a slide between two vowels, similar to that which would be observed in a diphthong. The following figures demonstrate the change in formant patterns generated when a /i/ to /a/ vowel transition is applied to the area function of an impedance mapped mesh. A linear constriction function is

used in Figure 5.14. Comparison with formant patterns generated using a 1D waveguide chain model (as in Figure 4.19) is included in the form of the dotted line. The transition is achieved with a linear interpolation between r^2 area functions, using random noise excitation from a point source at the centre of the glottal end of the tract.

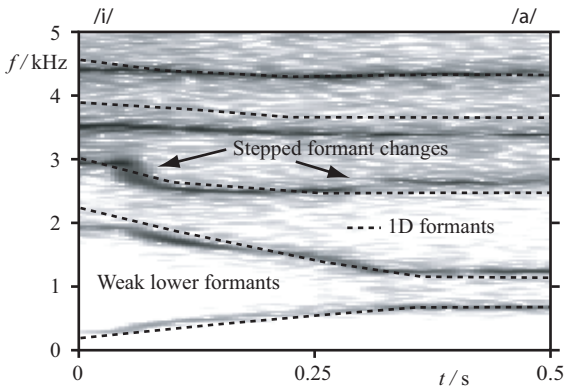


Figure 5.14: */i/ to /a/ vowel slide with linear impedance function map*

The overlayed dotted line highlights that the changes in formants are also very close to those generated using a 1D model. The lower two formants, however, are attenuated by approximately 12 dB compared to the third and fourth. Furthermore, the third formant can be seen to follow a slightly erratic path, giving stepped changes rather than the smooth transition required. Speech-like output generated using this model contains a buzzing quality where the glottal waveform is emerging with little of the important lower resonances imparted onto it. This is considered an artifact of the very narrow channel created with the linear impedance hills. The central Z_{min} lower impedance region has zero physical width (see Figure 5.2), and so has little of the desired effect as a direct propagational path.

The raised cosine impedance mapped model vowel slide formants are shown in Figure 5.15. Again, the frequencies appear at similar values to those generated with both the 1D and 2D widthwise models. Issues identified with the linear mapping function are no longer present. The relative strength and smoothness of transition of the formants is more in line with those observed in the 1D model.

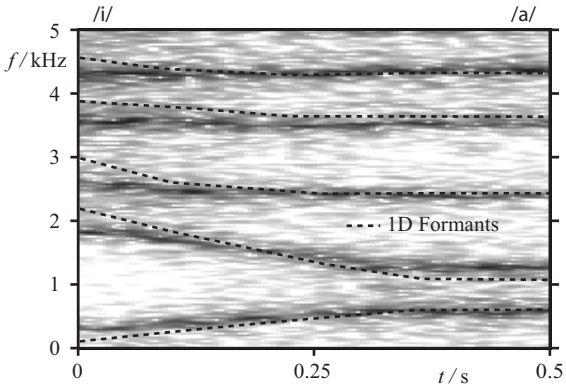


Figure 5.15: */i/ to /a/ vowel slide with raised cosine impedance function map*

An additional example is given in Figure 5.16 in the form of an /a/ to /ɛ/ raised cosine impedance mapped vowel slide.

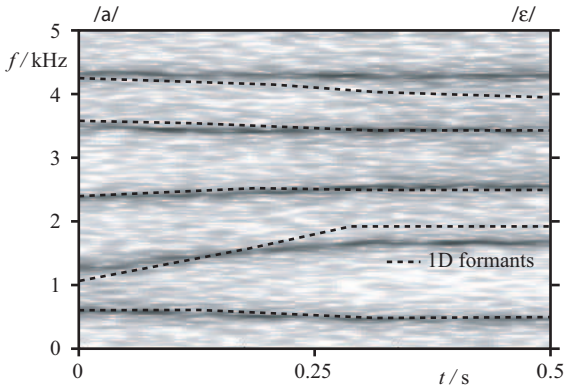


Figure 5.16: */a/ to /ɛ/ vowel slide with raised cosine impedance function map*

The wider central channel provided by the raised cosine method (Figure 5.3) provides an increased acoustic throughput, and therefore does not restrict the signal propagation as with the linear map. Figure 5.17 highlights this difference. Comparison is drawn between the raised cosine impedance function and the linear version (dotted line). The raised cosine function offers a wider central channel whilst still providing a smooth transition in the increased impedance effects towards the tract inner walls. Furthermore, vowel sounds generated with the raised cosine function are considered to be the most natural of the two simulations. As such, it is selected for use in further simulations.

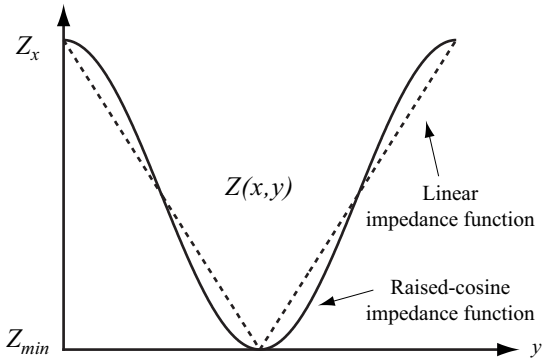


Figure 5.17: *Comparison of different constriction functions*

Clearly, the two functions offered here do not qualify as an exhaustive list of possible impedance variation shapes. It is worth reiterating that at the current state of development, the impedance mapping technique is at an experimental stage. Such arbitrary functions are chosen to demonstrate the possibilities of manipulating the waveguide impedances in this way.

5.7 Stable Articulations

Using the raised-cosine impedance function, there are no discontinuities audible in the resulting diphthong. This can be demonstrated with application of the LF glottal waveform as a plane source applied at the glottal end of the model during the slide between the two area functions. This is illustrated in Figure 5.18 where the output waveform after an area function change contains none of the high frequency discontinuities highlighted in the widthwise equivalent in Figure 4.22.

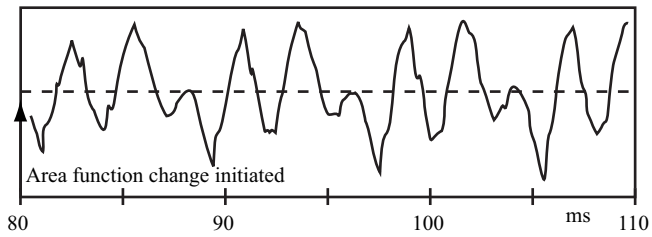


Figure 5.18: *No discontinuities in waveform at time of area function update*

5.7.1 Real-Time Operation

The *impedance mapping* techniques place no limitations on minimum width, as the mesh retains its rectangular shape throughout. Therefore a 2D system may be constructed with a waveguide size $d = 11$ mm, which gives a sampling frequency of $f_s = 44.1$ kHz, and given the $f_s/4$ restraints for the rectilinear mesh, a valid bandwidth of approximately 11 kHz. Such an arrangement employing a rectilinear mesh topology comprises 60 waveguide junctions. Exploiting this reduction in sampling frequency, software has been developed which demonstrates real-time 2D DWM vocal tract model vowel shape manipulation. The model allows for real-time user interaction using a mouse to bring about smooth vowel slides resulting in diphthongs and sharper area function changes affecting momentary constrictions to the airflow for simulation of plosive articulation.

Figure 5.19 shows example output from the software developed to simulate the 2D mesh vocal tract. In the simulation, voiced plosive articulation is modelled in the 2D impedance mapped /i/ vowel mesh using a mouse controlled slider which represents the area function at the lip end. The plosive is generated with a constriction to the tract as indicated by the high impedance Z_{stop} in Figure 5.4. The waveform has four distinctive points where a stop and release of pressure is generated as a direct result of the real-time constrictions made with the user interface. Combined with LF glottal excitation, each plosive part of the waveform is a synthesized version of the word *'bee'*. Other vowels and area function manipulations at different places along the tract may be used to simulate further voiced and non-voiced plosive articulations.

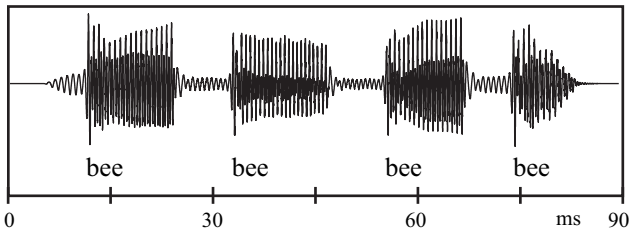


Figure 5.19: *Dynamic articulations using the real-time impedance mapped mesh model*

5.8 Conclusions

In this chapter, changes in waveguide impedance have been shown to provide a novel alternative method of area function application to the 2D DWM vocal tract model. It has been demonstrated that equivalent formant patterns to those generated with the well-established 1D waveguide chain model can be produced with a rectangular 2D mesh with an overlayed impedance map using the same area functions. Different approaches of defining the impedance map have been discussed. It was also noted that the relative size of the constrictions can be increased to enhance the extent to which the resulting formants move away from the equally spaced, neutral positions. Using the new method, stable, dynamic, real-time changes to the tract space modelled by the DWM were demonstrated in the form of synthesised diphthongs and plosive articulations.

Chapter 6

Summary and Analysis

6.1 The 2D DWM Vocal Tract Model

The well-established 1D waveguide vocal tract model was originally proposed by Kelly and Lochbaum [18] in 1962. The simplicity and intuitive manner in which the tract is sampled as a series adjoining cylinders makes it a good introduction into the wider field of physical modelling. As such, it is frequently presented in the surrounding literature as the archetype upon which more detailed acoustical simulations are built [10] [56] [46] [9] [57]. Developments, such as the inclusion of a nasal tract, lip radiation and wall losses have been used to synthesise the singing voice [107]. The use of fractional waveguides has been investigated as a method of making lengthwise changes to the tract shape [108] [110]. Further studies have been conducted into improving the wave propagation mechanism. Traditional waveguides can be substituted for conical equivalents that use scattering methodology derived from the spherical wave equation. This increases the accuracy of the model, giving higher-order area function approximation, but also adds to the computational load [108] [109].

The work contained in this thesis is an investigation into an increased dimensional representation in the Kelly-Lochbaum vocal tract model. Extension of the wave propagation mechanism into 2D has been presented as an alternative development to the basic 1D waveguide model over methods sim-

ulating enhanced order area function approximation. The digital waveguide mesh (DWM) physical model is typically used in virtual simulations of the acoustics of a room [13] [37] [15]. In this research it has been used to create a structure that represents the air cavity contained within the human vocal tract. A 2D DWM is formed such that lengthwise pressure wave propagation and reflection is sustained along the model between the glottis and lip ends. This is analogous to the planar wave propagation offered by the 1D system. The additional dimension in the 2D DWM is distributed across the tract, such that the simulated waveform within the mesh includes propagation and reflection in between the tract inner walls. In the same way that the 1D model neglects non-planar considerations, this 2D model forms a pseudo representation of the real acoustics of the tract. Modal interactions in the remaining dimension are not taken into account. However, the 2D model is intended as a proof-of-principle investigation. Factors that have been taken into consideration in extending the mesh across the tract are directly applicable to similar efforts in modelling the third dimension. This work, therefore, demonstrates the potential of such an augmented representation system, and some of the issues that might need to be addressed.

6.2 1D - 2D Model Comparison

To begin with, chapter 4 demonstrated how the modal resonances of a rectangle can be simulated with a 2D DWM. The waveguide structure was the same length as the average male vocal tract, 17.6 cm, with a width that is approximately equal to some of the larger openings observed in speech, 4 cm. All four reflecting boundaries were given a fully positive reflection. Comparison with the frequency peaks generated with a 1D waveguide model of the same length clearly show the additional modal patterns introduced with the extra dimension. Cross-axial modes begin to appear in the frequency response of the mesh from about 4.3 kHz upwards.

This is not a model of the vocal tract. A negative reflection should be set at the lip boundary, and tract widths of 4 cm are observed only in some

places along the tract, and only occasionally during speech. Nevertheless, the results do serve to highlight the additional acoustical properties that are introduced with the use of higher dimensionality in the modelling structure.

6.3 Area Function Data

The shape of the vocal tract that is held for production of each of the many speech sounds can be quantified in a 3D MRI scan. A large quantity of data is produced for each scan. Publication of such detailed area functions is often impractical, and so no 3D tract data was obtained for use in this work. Typically, the data is reduced to form a series of cross-sectional area values, each extracted from a plane that is perpendicular to the planar wave motion along the tract. This results in a shape description that is ideally suited to the 1D piecewise acoustic cylinder model. Axial asymmetries around the tract wall are disregarded, such that each area value represents a circular disc. The bend half way along the length of the tract is also neglected, such that it forms a straight tube. Such 1D area functions are readily available [76] [77] [78].

Vocal tract simulations presented in this work have used the 1D area functions to define the shape of the 2D mesh. This implies that the model will inherit the assumptions placed on circular cross-sectional area values and a straight tract. With shape information that contains greater detail, the higher dimensional model could be further extended, as it removes the requirements for such simplifications to be made. A 2D mesh could be constructed that takes more of, but not all, the axial asymmetries around the tract inner-wall into consideration. Similarly, a mesh could also incorporate the bend observed in the real tract. These effects were not investigated in this work.

Clearly, a 3D DWM vocal tract model constructed around full, 3D MRI data would provide highly accurate acoustical simulation. All shape-based approximations in the 1D and 2D models would not be necessary. Axial asymmetries and the tract bend would be completely facilitated. Furthermore, the 3D air cavity would allow for simulation of additional vocal conditions. For example, the tongue position held for production of the

lateral /l/ creates two air-channels, one either side of the tip. This scenario could be fully realised with a 3D DWM.

6.4 Widthwise Area Function Mapping

It was shown in Chapter 4 how the tract shape detail contained within the 1D area functions can be applied to the 2D model. The mesh is configured such that the width at each point is calculated as the diameter of the equivalent circular cross-section at each point along the 1D model. Changes to the tract shape happen proportionally to r, the radius of the corresponding 1D cylinder. It was discovered that with the mesh configured in this way, the resulting formants were shifted away from the neutral positions in the directions of those observed in the 1D model. However, the extent to which the formants had moved was less than expected. Vowels synthesised with the application of the LF glottal model were stated as having some of the audible qualities of the neutral vowel /ə/.

A variation on this method was also introduced that defines the width of the mesh as proportional to r^2. It was shown that this method may have some grounding in the angular volume summation that is used in defining the 3D space as a 2D mesh. However, no mathematical proof was offered for the manipulation of the area function. The changes have the effect of boosting the effects of the area function, such that the difference between large and small values is increased. Using enhanced area values for the width, the formant patterns were more in line with those from the 1D model. Furthermore, it is considered that the artificial vowel sounds can be described as bearing a strong resemblance to real world equivalents.

Disadvantages of the waveguide mesh vocal tract model were highlighted in Section 4.9. The DWM is a discrete modelling method and so only exists at finite spatial sampling instants. If the mesh shape is dynamically altered, the stepped movement from one sample location to another produces discontinuities in the resulting waveform. Furthermore, the limitations placed on a minimum channel width of two waveguides also

raise functionality issues. A higher resolution mesh with smaller waveguides will allow for a smaller minimum width channel, but this will involve a higher sampling frequency, and hence greater computational requirements. A vocal tract model that is limited to synthesis of static vowel sounds, with high processing demands and non real-time performance would not prove a particularly useful artificial speech generation tool.

6.4.1 Formant Bandwidths

In Section 4.8.2 it was demonstrated that the 2D mesh model offers sensitive bandwidth adjustment. Formant bandwidths contribute to the naturalness of vowel sounds. The ability to adjust synthesised bandwidth values towards those observed in natural speech will increase the power of a vocal tract model. In the 1D model, energy reflected back into the tract is largely governed by coefficients r_l and r_g. Results from the 2D model give little variation in bandwidth when r_g is used as a controlling parameter with wall losses set, equivalent to bandwidth variation in the traditional 1D model. In contrast, when the additional reflection coefficient r_w is used as a controlling parameter bandwidth values follow an approximately linear pattern of adjustment and hence optimum values can be achieved. The theoretically predicted values set in r_l and r_g are valid and can therefore remain fixed.

The simulated bandwidths do however, remain interrelated and currently cannot be individually tuned. Further research remains in the field of separating control of each of the three formant bandwidths, allowing fully optimized bandwidths for each formant of each vowel. This may be achieved by identifying which tract sections have a dominating influence on individual bandwidths and allowing for variable r_w values along the length of the tract.

6.4.2 Energy Losses

Investigations have been conducted into the frequency-dependent reflections in the tract at the lips in the 1D model [96] [97] [115]. Losses in the

tract walls have been accommodated along the length of the 1D frequency-domain circuit component tract model [83], [95], [96]. Implementation of tract losses in the 2D model in this work was directed towards simple frequency-independent reflections along the glottis, inner-wall and lip boundaries. Research into the application of the various forms of frequency dependent losses to the DWM tract model should increase the accuracy of synthesis. Considerations might involve how filter units could be employed to bring about these more accurate energy losses in the various junction boundary types, such as multi-port and variable impedance.

6.5 Impedance Area Function Mapping

Some of the disadvantages associated with the widthwise mapped mesh were addressed in Chapter 5. Attempts were made to develop a 2D mesh model which could accommodate the dynamic shape variations taking place within the vocal tract during speech. Inspiration was drawn from the archetypal 1D model, and the manner in which the area function changes are made to the impedance values of each waveguide. This technique was adapted to create a 2D impedance map that is applied to a mesh, which retains its rectangular structure throughout simulations. The area function changes therefore take place within the parameters of the waveguides, rather than in the shape of the structure. It was seen to generate similar formants peaks to those observed in the frequency response of the 1D model. However, the extent to which the simulated 2D formants matched the 1D benchmark was again increased with the use of an enhanced area function. Maps generated with impedance values that were inversely proportional to r^3 offered an improved synthesis over those based on an r^2 relationship. This conclusion was based on a greater perceptual correlation with the real-world vowel sounds.

6.5.1 Dynamic Operation

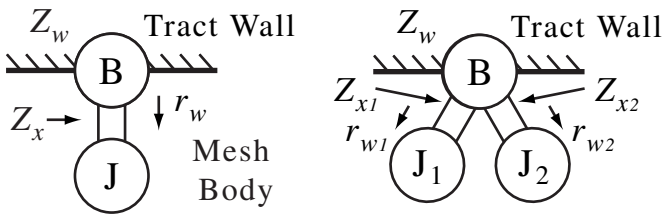
The impedance mapping technique allows for dynamic changes to be made to the resonant characteristics of the mesh without introducing discontinuities into the simulated waveform. Linear interpolations between area functions give rise to a steady transition between formant patterns. This is the first demonstration of a dynamically moving DWM, and as such it potentially opens up new areas of research into multidimensional representations of the vocal tract, as well as more general waveguide modelling applications.

The limitations placed on minimum width and high sampling frequencies are no longer of concern. These improvements in the model allowed for development of interactive real-time software which can be used to create basic articulatory synthesis. Mouse-controlled sliders on the application dialog box can be used to change the area function values applied to the mesh. Using these, sudden changes can be made to produce synthesis of realistic plosive sounds.

6.5.2 Mesh Topology and Boundary Considerations

The impedance map method introduces additional considerations in terms of tract inner wall boundary implementation. The three interdependent factors at a boundary junction are the reflection coefficient r_w, the wall material impedance Z_w, and the mesh body impedance defined as Z_x at each point along the length. This arrangement is illustrated in Figures 6.1(a) and 6.1(b) as the rectilinear and triangular mesh forms, respectively. The boundary junction B separates the mesh body junctions J and the tract wall.

As defined in the area function, and hence in the impedance map, Z_x varies along the edge of the mesh body. This implies that at least one of the other factors will also be nonuniform. The wall impedance can be kept constant while the reflection at each junction varies, or the wall boundary can be set to provide a constant r_w, with a varying Z_w. The notion of a variable wall impedance raises questions, as the tract is constructed of similar fleshy material along its length. It follows that it should be of constant impedance.



(a) Rectilinear 1-port (b) Triangular 2-port

Figure 6.1: *Mesh boundary configuration*

However, the formulation of the one-port boundary (Figure 6.1(a)) scattering equation requires that either r_w or both connecting impedances Z_x and Z_w must be known (this was demonstrated in Equations (2.82) - (2.85), where two admittance values either side of a discontinuity are substituted for a single reflection coefficient). Considering this notion raises the question of the physical values for the varying wall impedances. It is not clear how such data could be determined. Given that the model is based on manipulating the mesh body impedances, the definition of a constant Z_w would result in a largely variable reflection coefficients, and hence unpredictable energy losses in relation to the changes in Z_x. It was decided for this reason to specify a constant reflection coefficient at the value $r_w = 0.92$ obtained in the bandwidth simulations in Section 4.8.2. The different wall impedances along the length are therefore implicitly dealt with by each corresponding mesh body waveguide of Z_x and this uniform reflection.

This is not the case for the triangular topology mesh as it requires the use of multiple port boundary junctions along the edges. Such junctions were examined in Section 2.6.3 and an example of the two-port boundary is given in Figure 6.1(b). The scattering equation (2.113) defines each output in terms the reflection from that same port and the signal transmitted through from the other connection. The reflection coefficient seen at each port of the junction is different and depends on the relationship between all the impedances, including that of the boundary, in this case, Z_w. Because the multi-port boundary junction has different Z_x connections - Z_{x1} and Z_{x2} - as well as a Z_w, a constant reflection cannot be configured between the wall and

mesh body. In other words $r_{w1} \neq r_{w2}$. Moreover, there cannot be a uniform r_w along the wall length. It follows that some arbitrary value for Z_w must be defined, corresponding to approximate energy loss expectations in relation to the mesh body impedances at each junction.

A triangular mesh was constructed to implement the impedance mapped tract model in an attempt to examine the effects of the improved accuracy topology. However, for the reasons explained above, a stable structure was not successfully achieved. The different mesh body waveguide impedances at the boundary junction are determined from the map, as translated from the area function. A constant reflection cannot be arranged along the tract inner wall. Arbitrary wall impedance values were selected to match the non-uniform mesh body impedances. However, the reflections from each connection vary to a large extent. It is believed that waves incident upon the irregular wall boundary experienced anomalous reflections. Unpredictable results were obtained from the model. Further clarification of the role of the multi-port, different-impedance boundary junction should allow for formulation of a stable triangular mesh tract model of this type.

6.6 Unresolved Issues

In this work the exploration of how to translate the area function into the impedance map lead to two specific ambiguities; how to represent the 3D space as a 2D plane, and how to impart the impedance map across the mesh as an impedance function. These issues were not resolved, but were addressed with discussions on the increased power of r in the area functions, and in providing two functions to bring about the impedance transition across the mesh, respectively. Such ambiguities should be tackled in a continuation of this work, although at the current stage of development the approaches suggested here are sufficient for demonstrating the new techniques. Any computer based model will require many simplifications and approximations in order to meet the desired capabilities within reasonable computational restrictions. In this work such approximations are acceptable

for the proof-of-principle investigation, even if they deviate from a strict physically defined representation.

A further issue that has been raised concerns the use of excitation in the simulations. Results presented in Chapters 4 and 5 were generated with two different types of excitation; point-source random-noise input for analysis of the frequency response of the various meshes, and plane-source LF glottal input for demonstrating the vowel sound spectra. In many of the simulations, comparisons were drawn with the point-source noise excitation of the 2D DWM tracts and the equivalent 1D model. Excitation of the 1D model is equivalent to injection of an input across the whole of the tract surface area. Analysis of the 2D system using a point source input will therefore produce results from a different perspective to the 1D surface excitation. However, the formant patterns that were generated remain a valid analysis of the capabilities of the 2D tract model, given the stage of current development. A more physically related glottal input of the mesh tract would involve a plane or surface source excitation, such as that used for the vowel simulations.

6.7 Future Work

Further research relating to this work should consider the following issues.

Vocal Features

The central idea presented in this work is the development of a time-domain vocal simulator that removes some of the spatial sampling assumptions of the 1D model. Focus was placed on the propagational space contained within the vocal tract, and not on the interaction with adjoining regions, such as the nasal tract and subglottal regions. Such additional vocal features have been developed for the 1D model, and could be adapted for application in the 2D model. Construction of a 2D nasal cavity could be achieved using the same methods presented here. The two meshes could be combined at the velum by defining a side branch in the vocal tract using waveguide junctions with additional connections.

Mesh Topology

Further investigation of the use of different topology mesh structures should be undertaken. The increased propagational accuracy offered by the triangular and bilinearly deinterpolated meshes could prove beneficial to the resulting naturalness of synthesis. Furthermore, clarification of the scattering equations for triangular boundary junctions for use in impedance mapped models would be of use. This should lead to a dynamic real-time triangular mesh tract model.

Tract Energy Losses

Boundary implementation in this work has focused on simple reflection modelling. Future work should also consider how more accurate tract losses could be accommodated in the 2D model. Energy losses due to frictional, thermal and yielding wall effects could be incorporated along the lengths of the mesh as filter units. Consideration of how the frequency dependent reflection at the lip opening could be accommodated in the mesh should also increase the overall accuracy of the model.

Perceptual Tests

The 2D model has been presented here as a system that includes additional dimensional representation, and hence increased modal characteristics of the modelled space. It has not, however, been established as to whether these higher order, cross-tract acoustical properties are of importance to the perceived naturalness of synthesis. Perceptual tests could be used to determine the extent to which the additional resonant characteristics of the 2D model are of significance, and also how its resulting naturalness compares to the range of current 1D models.

Glottal Excitation

Glottal excitation used to simulate vowels was achieved with application of the LF waveform. A physical model of the vocal folds should be a more

accurate representation [83] [84] [85]. It would be of interest to investigate additional dimensionality in a glottal model, and how such a system could be connected with a multidimensional tract for interaction between the two.

Dynamic Modelling

The impedance mapping of area function was developed to address the problematic issues associated with the widthwise mapping alternative. Accurate formant patterns and realistic vowels can be synthesised, but the rectangular structure and impedance function act only to bring about similar resonant changes in the system. A direct equivalence or mathematical proof examining the relationship between impedance and widthwise mapping cannot be offered at this stage. Future work may involve a more formal definition of the techniques.

An exploration of alternative impedance functions across the mesh might also prove useful. For example, the impedance variations across the mesh do not necessarily need to be a strictly defined mathematical function. A more physically meaningful version of the technique might involve the use of moving impedance boundaries. The impedance map could be defined with two distinct regions; a lower Z_{air} through the centre of the tract and higher Z_{flesh} towards the edges of the rectangular mesh. A sinusoidal function similar to the one used in this work could be used to give a smooth transition between the two. These flesh-air boundaries could be configured with the ability to move inwards towards the tract centre to create a constriction, and away from each other to provide an opening, whilst retaining their form. The distance in between the two Z_{flesh} regions could be directly obtained from the area function and hence provide a more physically meaningful impedance map.

As additional work, it might also be beneficial to re-examine the manipulation of scattering junctions in the widthwise mapped mesh. Dynamic restructuring of waveguide junctions for movable boundaries were not successfully implemented in this work. This may be possible with further

research, possibly including fractional delay techniques [56], although at an increased computational load. The ability to make dynamic adjustments to the length of the mesh should also be investigated. This would increase modelling accuracy when forming tract shapes of varying lip protrusion.

Clarification of the Space Represented by the 2D Mesh

Results presented in this work showed that the formant synthesis offered by the 2D mesh was approximately equivalent to the 1D model. The perceptual match with real-world vowels was slightly improved with manipulation of the area function, such that values as applied to the mesh are raised to an additional power of r. This increased the extent to which the formants were shifted away from the neutral positions. This finding may have some grounding in theory, as indicated in Section 4.3.1, or it may just be a consequence of experimentation with the model. It might be of interest to consider formal definitions of the space that is actually modelled by the 2D mesh. Moreover, construction of the radial mesh tract model suggested in Section 4.3.1 may introduce further potential research directions for including the effects of the neglected third dimension in the 2D mesh.

A 3D DWM Vocal Tract Model

This proof-of-principle work acts to demonstrate the potential for construction of a full 3D model using a complete 3D scan of the tract shape. The use of full MRI scan data with complex-shape cross-sectional area data rather than the circular form used for 1D simulation should increase the naturalness of synthesis. With the inclusion of additional features, such as a nasal cavity, and frequency dependent energy losses, it is believed that highly accurate vocal synthesis could be achieved. A dynamic 3D model could also be developed. It is considered that the impedance mapping of area function presented in this work could be easily adapted for use in a 3D structure. A 17.6 cm long cuboid rectangular waveguide structure could have the area function set within the impedance of waveguides through each 2D slice

across the tract. Such a system would also lend itself well to the non-circular tract shapes found in the complex 3D scans. Such a DWM model would be well suited for use with techniques developed by Olav Engwall in 3D tract scanning and graphical representation [104].

Chapter 7

Conclusion

An analysis of the use of the 2D digital waveguide mesh (DWM) in modeling the acoustics of the human vocal tract has been presented. The research has been put forward as a technique for increasing the level of dimensional representation of the air-cavity within a vocal tract physical model beyond the well-established 1D method. Two methods for applying the area function data to the mesh have been demonstrated. The spectral results that were presented establish that accurate formant synthesis can be achieved using both techniques.

The first, widthwise mapping, was used to show that formant bandwidths follow a sensitive, linear response when the additional boundary reflection parameter is used to control energy losses. However, the static nature of the DWM structure is identified as a cause of discontinuities in the simulated waveform when its shape is dynamically altered. Widthwise mapping also defines a minimum channel width and requires an impractically high sampling frequency. The second area function application method presented, impedance mapping, uses a constant rectangular structure and tract shape changes are made within the waveguide impedances. This new technique allows for stable dynamic manipulations to be made to the cavity represented by the DWM. Moreover, the impedance-based method removes minimum channel and high sampling frequency requirements, allowing for a real-time response to be achieved.

The research demonstrates the potential for future work into a 3D model for highly accurate articulatory vocal synthesis. Simulation of additional factors, such as lip radiation, tract inner wall frequency dependent losses, a nasal cavity, tongue and jaw movement, would increase the potential of the model to produce natural sounding artificial speech sounds.

Appendix A

General Mathematics

A.1 The d'Alembert Solution to the 1D Wave Equation

The classic form of the 1D partial differential equation (PDE) where the wave variable u is dependant upon time t and the spatial coordinate x, with wave speed c, is

$$\frac{1}{c^2}\frac{\partial^2 u(x,t)}{\partial t^2} = \frac{\partial^2 u(x,t)}{\partial x^2} \tag{A.1}$$

The general solution of the 1D wave equation was proposed by the French mathematician Jean-le-Rond D'Alembert in 1747. The new variables $\xi(x,t) = x - ct$ and $\eta(x,t) = x + ct$ are introduced into the PDE (A.1) and the chain rule is used to express derivatives in terms of x and t as derivatives in terms of ξ and η [116]. The new function v is created

$$u(x,t) = v(\xi(x,t), \eta(x,t)) \tag{A.2}$$

Initially, first order differences are derived using the chain rule. Differentiating u with respect to x, noting that $\frac{\partial \xi}{\partial x} = 1$ and $\frac{\partial \eta}{\partial x} = 1$

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial \xi}\frac{\partial \xi}{\partial x} + \frac{\partial v}{\partial \eta}\frac{\partial \eta}{\partial x} = \frac{\partial v}{\partial \xi} + \frac{\partial v}{\partial \eta} \tag{A.3}$$

Similarly, differetiating u with respect to t, noting that $\frac{\partial \xi}{\partial t} = -c$ and $\frac{\partial \eta}{\partial t} = c$

$$\frac{\partial u}{\partial t} = \frac{\partial v}{\partial \xi}\frac{\partial \xi}{\partial t} + \frac{\partial v}{\partial \eta}\frac{\partial \eta}{\partial t} = -c\frac{\partial v}{\partial \xi} + c\frac{\partial v}{\partial \eta} \tag{A.4}$$

The second order differential with respect to x is

$$\begin{aligned} \frac{\partial^2 u}{\partial x^2} = \frac{\partial}{\partial x}\left(\frac{\partial v}{\partial \xi} + \frac{\partial v}{\partial \eta}\right) &= \frac{\partial^2 v}{\partial \xi^2}\frac{\partial \xi}{\partial x} + \frac{\partial^2 v}{\partial \eta \partial \xi}\frac{\partial \eta}{\partial x} + \frac{\partial^2 v}{\partial \xi \partial \eta}\frac{\partial \xi}{\partial x} + \frac{\partial^2 v}{\partial \eta^2}\frac{\partial \eta}{\partial x} \\ \implies \frac{\partial^2 u}{\partial x^2} &= \frac{\partial^2 v}{\partial \xi^2} + 2\frac{\partial^2 v}{\partial \xi \partial \eta} + \frac{\partial^2 v}{\partial \eta^2} \end{aligned} \tag{A.5}$$

Similarly, the second order differential with respect to t is

$$\begin{aligned} \frac{\partial^2 u}{\partial t^2} = \frac{\partial}{\partial t}\left(-c\frac{\partial v}{\partial \xi} + c\frac{\partial v}{\partial \eta}\right) &= -c\frac{\partial^2 v}{\partial \xi^2}\frac{\partial \xi}{\partial t} - c\frac{\partial^2 v}{\partial \eta \partial \xi}\frac{\partial \eta}{\partial t} + c\frac{\partial^2 v}{\partial \xi \partial \eta}\frac{\partial \xi}{\partial t} + c\frac{\partial^2 v}{\partial \eta^2}\frac{\partial \eta}{\partial t} \\ \implies \frac{\partial^2 u}{\partial t^2} &= c^2\frac{\partial^2 v}{\partial \xi^2} - 2c^2\frac{\partial^2 v}{\partial \xi \partial \eta} + c^2\frac{\partial^2 v}{\partial \eta^2} \end{aligned} \tag{A.6}$$

Substitution for the two second order differentials (A.5) and (A.6) into the wave equation (A.1) gives

$$\frac{\partial^2 v}{\partial \xi^2} - 2\frac{\partial^2 v}{\partial \xi \partial \eta} + \frac{\partial^2 v}{\partial \eta^2} = \frac{\partial^2 v}{\partial \xi^2} + 2\frac{\partial^2 v}{\partial \xi \partial \eta} + \frac{\partial^2 v}{\partial \eta^2} \tag{A.7}$$

Rearranging (A.7) reveals that

$$\frac{\partial^2 v}{\partial \xi \partial \eta} = 0 \tag{A.8}$$

This alternative form of the wave equation may be solved by direct integration. Firstly, with respect to ξ gives $\frac{\partial v}{\partial \eta} = g(\eta)$, where g is some arbitrary function of η. A further integration with respect to η gives

$$u(\xi, \eta) = F(\xi) + G(\eta) \tag{A.9}$$

Where F is some arbitrary function of ξ and $G(\eta) = \int g(\eta) d\eta$. Finally, replacing ξ and η by their expressions in terms of x and t

$$u(x,t) = F(x - ct) + G(x + ct) \tag{A.10}$$

This result describes the summation of two arbitrary waveforms F and G. In acoustical terms it provides a solution to the 1D wave equation for pressure variation p in terms of the summation of left and right going wave variables p_l and p_r at a point x in a 1D system

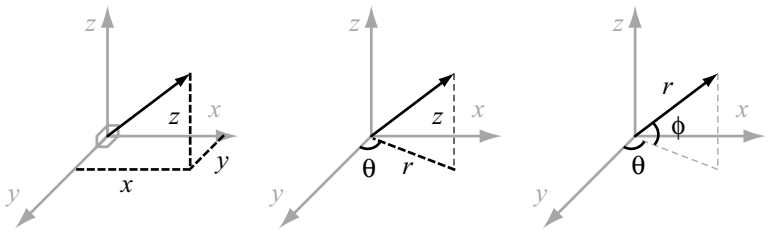
$$p(x,t) = p_l(x+ct) + p_r(x-ct) \tag{A.11}$$

A.2 Coordinate Systems

A general case for the wave equation in all coordinate systems can be defined such that the pressure p, moving at wavespeed c over time t is described by

$$\frac{1}{c^2}\frac{\partial^2 p(x,t)}{\partial t^2} = \nabla^2 p(x,t) \tag{A.12}$$

Where ∇^2, the Laplacian, is a differential operator used to characterise different coordinate systems. Figures A.1(a), A.1(b) and A.1(c) demonstrate the Cartesian, cylindrical polar and spherical polar coordinates systems, respectively.



(a) Cartesian x, y, z (b) Cylindrical r, θ, z (c) Spherical r, θ, ϕ

Figure A.1: *Coordinates systems*

In cartesian x, y, z coordinates the Laplacian is

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \tag{A.13}$$

In cylindrical polar coordinates r, θ and z the Laplacian is

$$\nabla^2 = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2}{\partial \theta^2} + \frac{\partial^2}{\partial z^2} \tag{A.14}$$

In spherical polar coordinates r, θ and ϕ, the Laplacian is

$$\nabla^2 = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2}{\partial\phi^2} \tag{A.15}$$

A.2.1 Coordinate Transformation

The volume in a three-dimensional space is often considered as an integral over the three coordinates. The integral of a function $f(x,y,z)$ can be evaluated in terms of a transformation to coordinates u,v and w [117].

$$\iiint f(x,y,z)\,dx\,dy\,dz = \iiint f[x(u,v,w),y(u,v,w),z(u,v,w)]J\,du\,dv\,dw \tag{A.16}$$

Where J is the Jacobian determinant

$$J = \left|\frac{\partial(x,y,z)}{\partial(u,v,w)}\right| = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} & \frac{\partial y}{\partial w} \\ \frac{\partial z}{\partial u} & \frac{\partial z}{\partial v} & \frac{\partial z}{\partial w} \end{vmatrix} \tag{A.17}$$

Using this, it is possible to translate from 3D cartesian coordinates x, y and z to cylindrical polar coordinates r, θ and z. From circular geometry, $x = r\cos\theta$ and $y = r\sin\theta$, and z is the same in both systems. The Jacobian determinant is

$$J = \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial\theta} & \frac{\partial x}{\partial z} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial\theta} & \frac{\partial y}{\partial z} \\ \frac{\partial z}{\partial r} & \frac{\partial z}{\partial\theta} & \frac{\partial z}{\partial z} \end{vmatrix} = \begin{vmatrix} \cos\theta & -r\sin\theta & 0 \\ \sin\theta & r\cos\theta & 0 \\ 0 & 0 & 1 \end{vmatrix} = r(\cos^2\theta + \sin^2\theta) = r \tag{A.18}$$

Therefore, the volume space V contained within a problem domain S can be obtained from integration in cartesian coordinates, or equivalently, integration in cylindrical polar coordinates adjusted by this additional r term.

$$V = \iiint S(x,y,z)\,dx\,dy\,dz = \iiint S(r,\theta,z)r\,dr\,d\theta\,dz \tag{A.19}$$

A.3 Separation of Variables

The variables in a partial differential equation (PDE) can be separated to form a number of ordinary differential equations (ODEs). This is the method used to isolate wave motion in a particular axis from others in a problem domain, such that it can be considered independently. For example, in classical acoustical tube simulations, vibrations are considered along the length, ignoring transverse effects. Here, a brief summary is given, following the techniques described in [43], [46] and [35].

A.3.1 Helmholtz Equation

The Helmholtz equation is the result of a separation of variables in a PDE, such that it becomes time-independent. It is possible to derive it with a substitution into the wave equation:

$$\frac{1}{c^2}\frac{\partial^2 p(x,t)}{\partial t^2} = \nabla^2 p(x,t) \tag{A.20}$$

Separating the spatial from the temporal dependencies, a solution of the form $p(x,t) = p(x)e^{-j\omega t}$ can be attempted.

$$\frac{1}{c^2}\frac{\partial^2}{\partial t^2}\left[p(x)e^{-j\omega t}\right] = \nabla^2\left[p(x)e^{-j\omega t}\right] \tag{A.21}$$

Noting that the left hand term is

$$\begin{aligned}\frac{1}{c^2}\frac{\partial^2}{\partial t^2}\left[p(x)e^{-j\omega t}\right] &= \frac{1}{c^2}j^2\omega^2 p(x)e^{-j\omega t} &\text{(A.22)}\\ &= -\frac{\omega^2}{c^2}p(x)e^{-j\omega t} &\text{(A.23)}\end{aligned}$$

Therefore (A.21) becomes

$$\nabla^2 p(x)e^{-j\omega t} + \frac{\omega^2}{c^2}p(x)e^{-j\omega t} = 0 \tag{A.24}$$

Finally, dividing through by $e^{-j\omega t}$ gives the classical form of the Helmholtz equation, where $k^2 = \omega^2/c^2$ relates to the wave number k and angular

frequency ω of the harmonic solution.

$$\nabla^2 p(x) + k^2 p(x) = 0 \tag{A.25}$$

Therefore harmonic solutions to the wave equation also satisfy the Helmholtz equation.

A.3.2 Cartesian Coordinates

The wave equation (A.12) for pressure p in 3D cartesian coordinates x, y, and z, is

$$\frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} = \frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} + \frac{\partial^2 p}{\partial z^2} \tag{A.26}$$

It can be stated that a separable solution exists as a product of four arbitrary functions $X(x)$, $Y(y)$, $Z(z)$ and $T(t)$ [43].

$$p(x,y,z,t) = X(x)Y(y)Z(z)T(t) \tag{A.27}$$

Substituting this into the wave equation A.26 and rearranging gives

$$\frac{1}{c^2}\frac{1}{T}\frac{\partial^2 T}{\partial t^2} = \frac{1}{X}\frac{\partial^2 X}{\partial x^2} + \frac{1}{Y}\frac{\partial^2 Y}{\partial y^2} + \frac{1}{Z}\frac{\partial^2 Z}{\partial z^2} \tag{A.28}$$

The left had side of (A.28) is dependent only on t, and the right hand side is dependent on x, y and z. It is therefore possible to separate the equations and equate them both to some arbitrary constant $-k^2$.

$$\frac{1}{c^2}\frac{1}{T}\frac{\partial^2 T}{\partial t^2} = -k^2 \tag{A.29}$$

$$\frac{1}{X}\frac{\partial^2 X}{\partial x^2} + \frac{1}{Y}\frac{\partial^2 Y}{\partial y^2} + \frac{1}{Z}\frac{\partial^2 Z}{\partial z^2} = -k^2 \tag{A.30}$$

Equation (A.30) can be further separated. The first term is dependent only on x, the second on y, and the third on z. Therefore

$$\frac{1}{X}\frac{\partial^2 X}{\partial x^2} = -k_x^2 \tag{A.31}$$
$$\frac{1}{Y}\frac{\partial^2 Y}{\partial y^2} = -k_y^2 \tag{A.32}$$
$$\frac{1}{Z}\frac{\partial^2 Z}{\partial z^2} = -k_z^2 \tag{A.33}$$

Where the relationship between them is

$$k_x^2 + k_y^2 + k_z^2 = k^2 = \frac{\omega^2}{c^2} \tag{A.34}$$

This separation constant $-k^2$ relates to the wave number k and angular frequency ω of the harmonic solutions to the wave equation. The four separate ordinary differential equations are now

$$\frac{\partial^2 T}{\partial t^2} + k^2 c^2 T = 0 \tag{A.35}$$
$$\frac{\partial^2 X}{\partial x^2} + k_x^2 X = 0 \tag{A.36}$$
$$\frac{\partial^2 Y}{\partial y^2} + k_y^2 Y = 0 \tag{A.37}$$
$$\frac{\partial^2 Z}{\partial z^2} + k_z^2 Z = 0 \tag{A.38}$$

Which are all forms of the Helmholtz equation A.25, solutions of which also satisfy the wave equation and can be solved independently from each other [43].

A.3.3 Cylindrical Polar Coordinates

The same techniques can be used to separate the modal variations across an acoustic tube, in r and θ terms from those relating to the length in z terms [43]. The wave equation in cylindrical coordinates is

$$\frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} = \frac{1}{r}\frac{\partial p}{\partial r}\left(r\frac{\partial p}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 p}{\partial \theta^2} + \frac{\partial^2 p}{\partial z^2} \tag{A.39}$$

A solution in terms of arbitrary functions $R(r)$, $\Theta(\theta)$ and $Z(z)$ is defined

$$p(r,z,\theta) = R(r)\Theta(\theta)Z(z) \tag{A.40}$$

Substitution of (A.40) into (A.39) gives

$$\frac{1}{c^2}\frac{1}{T}\frac{\partial^2 T}{\partial t^2} = \frac{1}{R}\frac{\partial^2 R}{\partial r^2} + \frac{1}{rR}\frac{\partial R}{\partial r} + \frac{1}{r^2\Theta}\frac{\partial^2 \Theta}{\partial \theta^2} + \frac{1}{Z}\frac{\partial^2 Z}{\partial z^2} \tag{A.41}$$

The following terms can be defined as

$$\frac{1}{c^2}\frac{1}{T}\frac{\partial^2 T}{\partial t^2} = -k^2 \tag{A.42}$$

$$\frac{1}{\Theta}\frac{\partial^2 \Theta}{\partial \theta^2} = -m^2 \tag{A.43}$$

$$\frac{1}{Z}\frac{\partial^2 Z}{\partial z^2} = -k_z^2 \tag{A.44}$$

Where an additional factor k_r is introduced such that

$$k_r^2 + k_z^2 = k^2 = \omega^2/c^2 \tag{A.45}$$

This leads to the wave equation for modal resonances with respect to r

$$\frac{\partial^2 R}{\partial r^2} + \frac{1}{r}\frac{\partial R}{\partial r} + \left(k_r^2 - \frac{m^2}{r^2}\right)R = 0 \tag{A.46}$$

This is a form of Bessel's equation of order m. Solutions can be found using the Bessel Function, which is described in greater detail in Section A.3.4. Solutions to the cylindrical wave equation in the r axis are a sum of Bessel functions of the first J_m and second Y_m kind.

$$R(r) = J_m(k_r r) + Y_m(k_r r) \tag{A.47}$$

A.3.4 The Bessel Function

A Bessel function provides solutions to the Bessel differential equation, which is of the form

$$x^2\frac{\partial^2 y}{\partial x^2} + x\frac{\partial y}{\partial x} + (x^2 - m^2)y = 0 \tag{A.48}$$

This second order differential has two linearly independent solutions, a Bessel function of the first kind $J_m(x)$, and of the second kind $Y_m(x)$, weighted by coefficients c_1 and c_2, respectively

$$y(x) = c_1 J_m(x) + c_2 Y_m(x) \tag{A.49}$$

In general a Bessel function of the first kind and order m is expressed as

$$J_m(x) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!\Gamma(n+m+1)} \left(\frac{x}{2}\right)^{2n+m} \tag{A.50}$$

The Bessel function of the first kind function evaluated with order $m = 0$ (black line), $m = 1$ (grey line) and $m = 2$ (dotted line) is illustrated in Figure A.2.

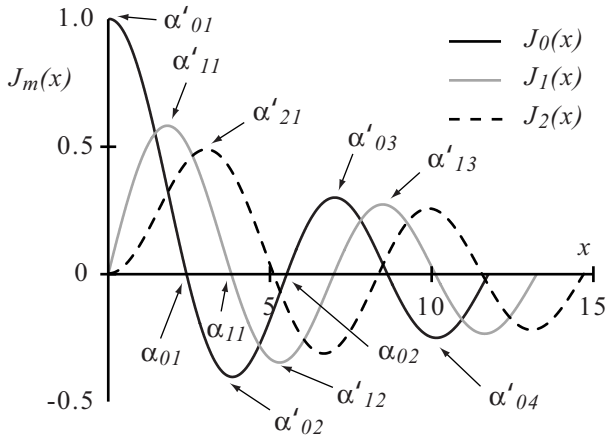


Figure A.2: *Bessel function of the first kind - order zero $J_0(x)$, one $J_1(x)$ and two $J_2(x)$*

The roots of the Bessel function are included on the diagram, such that α_{mn} is the x value where the nth instance of $J_m(x) = 0$ occurs - the zero crossings. Similarly, α'_{mn} is the x value where the nth instance of $\frac{dJ_m(x)}{dx} = 0$ occurs - the zero gradients. These roots are useful in determining the frequency of a Bessel function standing wave across the radius axis r of an acoustic tube. At the rigid reflecting boundary inside such a tube of radius a, standing waves form such that pressure components have zero gradient. From the diagram it can be seen that the first three allowable standing waves from the tube centre

at $r = 0$ to the tube edge $r = a$, in order of increasing frequency, are:

- From $x = 0$ to $\alpha'_{11} = 1.841$ (at $r = a$)
- From $x = 0$ to $\alpha'_{21} = 3.054$ (at $r = a$)
- From $x = 0$ to $\alpha'_{02} = 3.832$ (at $r = a$)

Appendix B

Phonology

B.1 Phonetic Alphabets

All phonemes from all languages produced using the human voice are identified, catagorised and represented with a unique symbol within the International Phonetic Alphabet (IPA) [74]. Many schemes exist to provide a translation for each symbol into a computer character for reproduction in written work. Two widely used methods are the numerically based Unicode system [118] and the Speech Assessment Methods Phonetic Alphabet (SAMPA) [119] which uses ASCII codes where Unicode is not available or not appropriate. The following tables outline the IPA symbols (generated using the LaTeXpackage `tipa`) commonly used in English and their corresponding SAMPA equivalent. Descriptions relating to the oral configuration held for utterance and example words are also given.

IPA	SAMPA	Description	Example Words
a	a	central open	f**a**ther, b**a**rd
i	i	front close unrounded	s**ee**, b**e**ad
ɪ	I	front close unrounded	c**i**ty, b**i**t
ɛ	E	front half open unrounded	g**e**t, b**e**d
ɜ	3	front half open unrounded	f**u**rry, b**I**rd
æ	{	front open unrounded	c**a**t, b**a**d
ʌ	V	back half open unrounded	r**u**n, b**u**t
ɔ	O	back half open rounded	l**a**w, b**ou**ght
ʊ	U	back close rounded	p**u**t, b**oo**k
u	u	back close rounded	s**oo**n, b**oo**ed
ɒ	Q	back open rounded	n**o**t, b**o**d
ə	@	central neutral unrounded	**a**bout, winn**e**r

Table B.1: *Vowel catagorisation (English) - IPA and SAMPA symbols*

IPA	SAMPA	Example Word
əʊ	@U	hope
aʊ	aU	house
aɪ	aI	kite
eɪ	eI	same
ju	ju	few
ɔɪ	OI	join
ɪə	I@	fear
ɛə	E@	hair
ʊə	U	poor

Table B.2: *Diphthong catagorisation (English) - IPA and SAMPA symbols*

IPA	SAMPA	Description	Example Words
b	b	bilabial plosive	**b**ut, we**b**
d	d	alveolar plosive	**d**o, o**dd**
g	g	velar plosive	**g**o, **g**et, be**g**
v	v	labiodental fricative	**v**oice, ha**v**e
ð	D	dental fricative	**th**is, brea**th**e
z	z	alveolar fricative	**z**oo, ro**s**e
ʒ	Z	postalveolar fricative	plea**s**ure, bei**g**e
m	m	bilabial nasal	**m**an, ha**m**
n	n	alveolar nasal	**n**o, ti**n**
ŋ	N	velar nasal	si**ng**er, ri**ng**
l	l	alveolar lateral approximant	**l**eft
ɫ	5	velarised alveolar lateral approximant	mi**l**k, be**ll**
ɹ	r\	alveolar approximant	**r**un, ve**r**y
w	w	labial-velar approximant	**w**e
j	j	palatal approximant	**y**es

Table B.3: *Voiced consonant catagorisation (English) - IPA and SAMPA symbols*

IPA	SAMPA	Description	Example Words
p	p	bilabial plosive	**p**en, ti**p**
t	t	alveolar plosive	**t**wo, be**t**
ʧ	tS	voiceless postalveolar fricative	**ch**air, na**t**ure, tea**ch**
k	k	velar plosive	**c**at, s**k**in
f	f	labiodental fricative	**f**ool, enou**gh**, lea**f**
θ	T	dental fricative	**th**ing, wi**th**
s	s	alveolar fricative	**s**ee, **c**ity, pa**ss**
ʃ	S	postalveolar fricative	**sh**e, **s**ure, lea**sh**
h	h	glottal approximant	**h**am

Table B.4: *Voiceless consonant catagorisation (English) - IPA and SAMPA symbols*

References

[1] J. Mullen, D. M. Howard, and D. T. Murphy. Real-time dynamic articulations in the 2D waveguide mesh vocal tract model. In *IEEE Transactions on Audio, Speech and Language Processing*, volume 15(2), pages 577–585, Feb 2007.

[2] J. Mullen, D. M. Howard, and D. T. Murphy. Waveguide physical modeling of vocal tract acoustics: Flexible formant bandwidth control from increased model dimensionality. In *IEEE Transactions on Speech and Audio Processing*, volume 14(3), pages 964–971, May 2006.

[3] J. Mullen, D. M. Howard, and D. T. Murphy. Acoustical simulations of the human vocal tract using the 1D and 2D digital waveguide software model. In *Proc. 7th Int. Conf. on Digital Audio Effects (DAFx-04)*, pages 311–314, Naples, Italy, 2004.

[4] J. Mullen, D. M. Howard, and D. T. Murphy. Digital waveguide mesh modelling of the vocal tract acoustics. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 119–122, New Paltz, New York, USA, 2003.

[5] P. A. Taylor, A. Black, and R. Caley. The architecture of the festival speech synthesis system. In *Proc. of the 3rd ESCA Workshop in Speech Synthesis*, pages 147–151, Jenolan Caves, Australia, 1998.

[6] T. Dutoit. High-quality text-to-speech synthesis: an overview. In *Journal of Electrical and Electronics Engineering, Australia: Special Issue on Speech Recognition and Synthesis*, volume 17(1), pages 25–37, 1998.

[7] M. N. O. Sadiku. *Numerical Techniques in Electromagnetics*. CRC Pr Llc, 2001.

[8] T. J. Chung. *Computational Fluid Dynamics*. Cambridge University Press, 2002.

[9] S. Bilbao. *Wave and Scattering Methods for the Numerical Integration of Partial Differential Equations*. PhD thesis, Stanford University, USA, 2001.

[10] J. O. Smith III. *Physical Audio Signal Processing: Digital Waveguide Modeling of Musical Instruments and Audio Effects, Draft Version*. Web

published at `www-ccrma.stanford.edu/~jos/waveguide/pasp`, Stanford University, USA, August 2004.

[11] J. O. Smith III. Efficient simulation of the reed-bore and bow-string mechanisms. In *Proc. International Computer Music Conference*, pages 275–280, The Hague, Netherlands, 1986.

[12] S. A. Van Duyne and J. O. Smith III. Physical modeling with the 2D digital waveguide mesh. In *Proc. International Computer Music Conference*, pages 40–47, Tokyo, Japan, 1993.

[13] S. A. Van Duyne and J. O. Smith III. The 3D tetrahedral digital waveguide mesh with musical applications. In *Proc. International Computer Music Conference*, pages 9–16, Hong Kong, 1996.

[14] J. Laird. *The Physical Modelling Of Drums Using Digital Waveguides*. PhD thesis, Bristol University, 2001.

[15] M. J. Beeson and D. T. Murphy. Roomweaver: A digital waveguide mesh based room acoustics research tool. In *Proc. 7th Int. Conf. on Digital Audio Effects (DAFx-04)*, pages 268–273, 2004.

[16] U. Stephenson. Comparison of the image source method & the particle simulation method. In *Applied Acoustics*, volume 29, pages 25–72, 1990.

[17] A. Krokstad, S. Strøm, and S. Srøsdal. Calculating the acoustical room response by use of a ray tracing technique. In *Journal of Sound and Vibration*, volume 8(1), pages 118–125, 1968.

[18] J. L. Kelly and C. C. Lochbaum. Speech synthesis. In *Proc. Fourth International Congress on Acoustics*, pages 1–4, Copenhagen, Denmark, 1962.

[19] C. Lu, T. Nakai, and H. Suzuki. Finite element simulation of sound transmission in vocal tract. In *Journal of the Acoustical Society of Japan*, volume 14(2), pages 63–72, 1993.

[20] T. Niikawa, M. Matsumura, T. Tachimura, and T. Wada. Modeling of a speech production system based on mri measurement of three-dimensional vocal tract shapes during fricative consonant phonation. In *6th International Conference on Spoken Language Processing (ICSLP)*, volume 2, pages 174–177, 2000.

[21] S. El-Masri, X. Pelorson, P. Saguet, and P. Badin. Vocal tract acoustics using the tlm method. In *Proc. International Conference on Speech and Language Processing*, volume 2, pages 953–956, Philadelphia, USA, 1996.

[22] R. Sproat. *Multilingual Text-To-Speech Synthesis: The Bell Labs Approach*. Kluwer Academic, 2001.

[23] P. R. Cook. Singing voice synthesis: History, current work and future directions. In *Computer Music Journal*, volume 20(3), pages 38–46, 1996.

[24] A. Horner, J. Beauchamp, and L. Haken. Methods for multiple wavetable synthesis of musical instrument tones. In *Journal of the Audio Engineering Society*, volume 41(5), pages 336–356, 1993.

[25] J. Chowning. The synthesis of complex audio spectra by means of frequency modulation. In *Journal of the Audio Engineering Society*, volume 21(7), 1973.

[26] R. Rabenstein and L. Trautmann. Digital sound synthesis by physical modelling. In *Symposium on Image and Signal Processing and Analysis (ISPA)*, Pula, Croatia, 2001.

[27] S. Lehman. *Physical Modeling Synthesis*. Harmony Central, `www.harmony-central.com/Synth/Articles/Physical_Modeling`, 1996.

[28] J. O. Smith III. Physical modeling synthesis update. In *Computer Music Journal*, volume 20(2), 1996.

[29] L. Hiller and P. Ruiz. Synthesizing musical sounds by solving the wave equation for vibrating objects: Part 1. In *Journal of the Audio Engineering Society*, volume 19, page 462470, 1971.

[30] J. L. Florens and C. Cadoz. *Representations of Musical Signals*, chapter The physical model: modeling and simulating the instruments universe, pages 227–268. MIT Press,Cambridge, MA, 1991.

[31] K. Karplus and A. Strong. Digital synthesis of plucked string and drum timbres. In *Computer Music Journal*, volume 7(2), pages 43–55, 1983.

[32] D. A. Jaffe and J. O. Smith III. Extensions of the karplus-strong plucked string algorithm. In *Computer Music Journal*, volume 7(2), pages 56–69, 1983.

[33] J. O. Smith III. Physical modeling using digital waveguides. In *Computer Music Journal*, volume 16(4), pages 74–91, 1992.

[34] V. Välimäki. Plucked string models: from the karplus-strong algorithm to digital waveguides and beyond. In *Computer Music Journal*, volume 22(3), pages 17–32, 1998.

[35] G. P. Scavone. *An Acoustic Analysis of Single-Reed Woodwind Instruments with an Emphasis on Design and Performance Issues and Digital Waveguide Modeling Techniques*. PhD thesis, Stanford University, USA, 1997.

[36] P. R. Cook. *Real Sound Synthesis for Interactive Applications*. A K Peters, 2002.

[37] J. Huopaniemi, L. Savioja, and M. Karjalainen. Modelling of reflections and air absorption in acoustical spaces a digital filter design approach. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, New York, USA, 1997.

[38] H. F. Olson. *Acoustical Engineering*. Van Nostrand, 1957.

[39] E. W. Weisstein. *Speed of Sound*. Eric Weisstein's World of Physics, `scienceworld.wolfram.com/physics/SpeedofSound`, 2006.

[40] T. D. Rossing. *The Science of Sound*. Addison-Wesley, 1989.

[41] D. M. Howard and J. Angus. *Acoustics and Psychoacoustics*. Focal Press, 1996.

[42] H. J. Pain. *The Physics and Vibrations of Waves*. John Wiley and Sons Ltd., 3rd edition, 1983.

[43] D. T. Blackstock. *Fundamentals of Physical Acoustics*. Wiley-Interscience, 2000.

[44] C. R. Nave. *HyperPhysics: Reflection of Sound*. Department of Physics and Astronomy, Georgia State University, `http://hyperphysics.phy-astr.gsu.edu/Hbase/hph.html`, 2000.

[45] Lord Rayleigh. *The Theory of Sound*. MacMillan Company, 1896.

[46] F. Avanzini. *Computational Issues in Physically-Based Sound Models*. PhD thesis, Dept. of Computer Science and Electronics, University of Padova, Italy, 2001.

[47] J. Calvert. *Cylindrical Coordinates - Problems with Axial Symmetry*. `http://www.du.edu/~jcalvert/math/cylcoord.htm`, 2000.

[48] F. E. Relton. *Applied Bessel Functions*. Blackie and Son, London, 1946.

[49] T. Smyth, J. Abel, and J. O. Smith III. The feathered clarinet reed. In *International Conference on Digital Audio Effects (DAFx-04)*, pages 95–100, Naples, Italy, 2004.

[50] B. Krach, S. Petrausch, and R. Rabenstein. Digital sound synthesis of brass instruments by physical modelling. In *International Conference on Digital Audio Effects (DAFx-04)*, pages 101–106, Naples, Italy, 2004.

[51] S. Adachi and M. Sato. Trumpet sound simulation using a two-dimensional lip vibration model. In *Journal of the Acoustical Society of America*, volume 99, pages 1200–1209, 1996.

[52] P. M. Morse and K. U. Ingard. *Theoretical Acoustics*. New York: McGraw-Hill, 1968.

[53] D. M. Howard, S. Rimell, and A. Hunt. Cymatic: A tactile controlled physical modelling instrument. In *6th Int. Conference on Digital Audio Effects (DAFx-03)*, London, UK, 2003.

[54] M. Karjalainen and C. Erkut. Digital waveguides versus finite difference structures: Equivalence and mixed modelling. In *EURASIP Journal on Applied Signal Processing*, volume 2004(7), pages 978–989, New Paltz, New York, USA, 2004.

[55] E. W. Weisstein. *Spherical Wave*. Eric Weisstein's World of Physics, `scienceworld.wolfram.com/physics/SphericalWave`, 2006.

[56] V. Välimäki. *Discrete-Time Modeling of Acoustic Tubes Using Fractional Delay Filters*. PhD thesis, Laboratory of Acoustics and Audio Signal Processing, Faculty of Electrical Engineering, Helsinki University of Technology, Finland, 1995.

[57] M. van Walstijn. *Discrete-Time Modelling of Brass and Reed Woodwind Instruments with Application to Musical Sound Synthesis*. PhD thesis, University of Edinburgh, UK, 2002.

[58] D. T. Murphy. *Digital Waveguide Mesh Topologies In Room Acoustics Modelling*. PhD thesis, University Of York, UK, 2000.

[59] V. Välimäki and M. Karjalainen. Digital waveguide modeling of wind instrument bores constructed of truncated cones. In *International Computer Music Conference*, pages 423–430, 1994.

[60] D. P. Berners. *Acoustics and Signal Processing Techniques for Physical Modeling of Brass Instruments*. PhD thesis, Stanford University, USA, 1999.

[61] D. P. Berners and J. O. Smith III. On the use of schrödingers equation in the analytic determination of horn reflectance. In *International Computer Music Conference (ICMC)*, pages 419–422, 1994.

[62] L. Savioja and T. Lokki. Digital waveguide mesh for room acoustic modeling. In *ACM SIGGRAPH and Eurographics Campfire: Conference on Acoustic Rendering for Virtual Environments*, SnowBird, Utah, USA, 2001.

[63] J. Strikwerda. *Finite Difference Schemes and Partial Differential Equations*. Chapman & Hall, New York, NY, 1989.

[64] F. Fontana and D. Rocchesso. Signal-theoretic characterization of waveguide mesh geometries for models of two dimensional wave propagation in elastic media. In *IEEE Transactions on Speech and Audio Processing*, volume 9(2), pages 152–161, 2001.

[65] L. Savioja. *Modeling Techniques for Virtual Acoustics*. PhD thesis, Helsinki University of Technology, Finland, 1999.

[66] L. Savioja and V. Välimäki. Reducing the dispersion error in the digital waveguide mesh using interpolation and frequency warping techniques. In *IEEE Transactions on Speech and Audio Processing*, volume 8(2), pages 184–193, 2000.

[67] L. Savioja and V. Välimäki. The bilinearly deinterpolated waveguide mesh. In *IEEE Nordic Signal Processing Symposium (NORSIG'96)*, pages 443–446, Espoo, Finland, 1996.

[68] V. Välimäki and L. Savioja. Interpolated and warped 2d digital waveguide mesh algorithms. In *International Conference on Digital Audio Effects (DAFx-00)*, pages 201–206, Verona, Italy, 2000.

[69] M. Aird, J. Laird, and J. Fitch. Modelling a drum by interfacing 2D and 3D waveguide meshes. In *Proc. of the International Computer Music Conference*, pages 82–85, Berlin, Germany, 2000.

[70] D. T. Murphy and J. Mullen. Digital waveguide mesh modelling of room acoustics: Improved anechoic boundaries. In *5th International Conference on Digital Audio Effects (DAFx-02)*, pages 163–168, Hamburg, Germany, 2002.

[71] A. Kelloniemi. Improved adjustable boundary condition for the 2-d digital waveguide mesh. In *Proc. 8th Int. Conf. on Digital Audio Effects (DAFx-05)*, Madrid, Spain, 2005.

[72] A. Kelloniemi, L. Savioja, and V. Välimäki. Spatial filter-based absorbing boundary for the 2d digital waveguide mesh. In *IEEE Signal Processing Letters*, volume 12(2), pages 126–129, 2005.

[73] J. O. Smith. *Applications of Digital Signal Processing to Audio and Acoustics*, chapter Principles of Digital Waveguide Models of Musical Instruments, pages 417–466. MIT Press,Cambridge, MA, 1998.

[74] The International Phonetic Association. *The Handbook of the International Phonetic Association*. Cambridge University Press, 1999.

[75] R. Lederman. *Vocal Focus*. `www.vocalfocus.com/vocal-anatomy`, 2005.

[76] G. Fant. *Acoustic Theory of Speech Production*. Mouton, The Hague, 1960.

[77] B. H. Story, I. R. Titze, and E. A. Hoffman. Vocal tract area functions from magnetic resonance imaging. In *Journal of the Acoustical Society of America*, volume 100(1), pages 537–554, 1996.

[78] C.D. Gray, D.M. Campbell, and C.A. Greated. Acoustic pulse reflectometry for vocal tract measurements. In *Proc. of the Stockholm Music Acoustics Conference (SMAC)*, Stockholm, Sweden, 2003.

[79] D. G. Childers. *Speech Processing and Synthesis Toolboxes*. John Wiley and Sons Inc., 2000.

[80] G. J. Borden and K. S. Harris. *Speech Science Primer: Physiology, Acoustics and Perception*. Williams & Wilkens, Baltimore, 1980.

[81] S. Maeda. The role of the sinus cavities in the production of nasal vowels. In *Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages 911–914, 1982.

[82] O. Fujimura. Analysis of nasal consonants. In *Journal of the Acoustical Society of Japan*, volume 34, page 18651875, 1962.

[83] K. Ishizaka and J. Flanagan. Synthesis of voiced sounds from a two-mass model of the vocal folds. In *Bell System Technical Journal*, volume 51, pages 1233–1268, 1972.

[84] B. H. Story and I. R. Titze. Voice simulation with a bodycover model of the vocal folds. In *Journal of the Acoustical Society of America*, volume 97(2), pages 1249–1260, 1995.

[85] I. R. Titze. The human vocal cords: A mathematical model, part I. In *Phonetica*, volume 28, pages 129–170, 1973.

[86] I. R. Titze. The human vocal cords: A mathematical model, part ii. In *Phonetica*, volume 29, pages 1–21, 1974.

[87] A. E. Rosenberg. Effect of glottal pulse shape on the quality of natural vowels. In *Journal of the Acoustical Society of America*, volume 49(2), pages 583–590, 1970.

[88] D. H. Klatt. Software for a cascade/parallel formant synthesiser. In *Journal of the Acoustical Society of America*, volume 67(3), pages 971–995, 1980.

[89] G. Fant, J. Liljencrants, and Q. Lin. A four parameter model of glottal flow. In *Quarterly Progress Report*, volume 4, Speech Transmission Laboratory, Royal Institute of Technology, Stockholm, Sweden, 1986.

[90] N. Henrich, C. dAlessandro, and B. Doval. Spectral correlates of voice open quotient and glottal flow asymmetry: theory, limits and experimental data. In *Proceedings of EUROSPEECH*, pages 47–50, Aalborg, Denmark, 2001.

[91] C. Gobl. *The Voice Source in Speech Communication*. PhD thesis, Royal Institute of Technology, Stockholm, Sweden, 2003.

[92] J. Liljencrants. *Speech Synthesis with a Reflection-Type Line Analogue*. PhD thesis, Royal Institute of Technology, Stockholm, Sweden, 1985.

[93] M. Kob. *Physical Modeling of the Singing Voice*. PhD thesis, Aachen University of Technology, Germany, 2002.

[94] L. R. Rabiner and R. W. Schafer. *Digital Processing of Speech Signals*. Englewood Cliffs, NJ: Prentice-Hall, 1978.

[95] S. Maeda. A digital simulation method of the vocal-tract system. In *Speech Communication*, volume 1, pages 199–229, 1982.

[96] J. L. Flanagan. *Speech Analysis, Synthesis and Perception*. Springer-Verlag, New York, 1972.

[97] P. Badin and G. Fant. Notes on vocal tract computation. In *Quarterly Progress Report*, volume 2-3, Speech Transmission Laboratory, Royal Institute of Technology, Stockholm, Sweden, 1984.

[98] J. N. Holmes. Formant synthesizers: Cascade or parallel? In *Speech Communication*, volume 2, pages 251–273, 1983.

[99] J. Holmes and W. Holmes. *Speech Synthesis and Recognition*. Taylor and Francis, London and New York, 2nd edition, 2001.

[100] A. Hunt and A. Black. Unit selection in a concatenative speech synthesis system using a large speech database. In *Proc. of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 373–376, Atlanta, Georgia, 1996.

[101] P. Birkholtz and D. Jackel. A three-dimensional model of the vocal tract for speech synthesis. In *Proc. of the 15th International Congress of Phonetic Sciences (ICPhS)*, pages 2597–2600, Barcelona, Spain, 2003.

[102] M. M. Sondhi. Resonances of a bent vocal tract. In *Journal of the Acoustical Society of America*, volume 79(4), pages 1113–1116, 1986.

[103] M. M. Sondhi and J. Schroeter. A hybrid time-frequency domain articulatory speech synthesizer. In *IEEE Transactions on Acoustics, Speech and Signal Processing*, volume 35(7), pages 955–967, 1987.

[104] O. Engwall. *Tongue Talking - Studies in Intraoral Speech Synthesis*. PhD thesis, Royal Institute of Technology, Stockholm, Sweden, 2002.

[105] H. W. Strube. Time-varying wave filters for modeling analogue systems. In *IEEE Transactions on Acoustics, Speech and Signal Processing (ASSP)*, volume 30, page 864868, 1982.

[106] B. H. Story. *Physiologically-Based Speech Simulation with an Enhanced Wave-Reflection Model of the Vocal Tract*. PhD thesis, University of Iowa, USA, 1995.

[107] P. R. Cook. *Identification of Control Parameters in an Articulatory Vocal Tract Model with Applications to the Synthesis of Singing*. PhD thesis, Stanford University, USA, 1991.

[108] V. Välimäki and M. Karjalainen. Improving the kelly-lochbaum vocal tract model using conical tube sections and fractional delay filtering techniques. In *Proc. Int. Conf. on Spoken Language Processing (ICSLP)*, pages 615–618, Yokohama, Japan, 1994.

[109] H. W. Strube. Are conical segments useful for vocal-tract simulation? In *Jounal of the Acoustical Society of America*, volume 114(6), pages 3028–3031, 2003.

[110] I. R. Titze. Vocal tract modeling: Implementation of continuous length variations in a half-sample delay kelly-lochbaum model. In *Proc. IEEE Intl. Symposium on Signal Processing and Information Technology*, Darmstadt, Germany, 2003.

[111] H. Matsuzaki, N. Miki, N. Nagai, T. Hirohku, and Y. Ogawa. 3d fem analysis of vocal tract model of elliptic tube with inhomogenous-wall impedance. In *3rd International Conference on Spoken Language Processing (ICSLP)*, pages 635–638, 1994.

[112] D. G. Childers and K. Wu. Gender recognition from speech. part ii: Fine analysis. In *Journal of the Acoustical Society of America*, volume 90(4), pages 1841–1856, 1991.

[113] J. Prosise. *Programming Windows with MFC*. Microsoft Press, 1999.

[114] OpenGL Architecture Review Board. *OpenGL Programming Guide: The Official Guide to Learning OpenGL*. Addison Wesley Professional, 1999.

[115] U. K. Laine. Modelling of lip radiation impedance in z-domain. In *Proc. Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, volume 3, pages 1992–1995, Paris, France, 1982.

[116] G. Bekefi and A. H. Barrett. *Electromagnetic Vibrations, Waves, and Radiation*. MIT Press, Cambridge, MA, 1987.

[117] H. Jeffreys and B. S. Jeffreys. *Methods of Mathematical Physics*, chapter Change of Variable in an Integral, pages 32–33. Cambridge University Press, 1988.

[118] The Unicode Consortium. *The Unicode Standard*. Addison-Wesley Professional, 2003.

[119] J. C. Wells. *Handbook of Standards and Resources for Spoken Language Systems*. Mouton de Gruyter, Berlin and New York, 1997.