VOCAL TRACT ACOUSTIC MEASUREMENTS AND THEIR APPLICATION TO ARTICULATORY MODELLING

Billy Hogan BE

Submitted for the degree of Master of Engineering (M.Eng.),
23/08/1996

Supervised by Dr Ronan Scaife
School of Electronic Engineering
Dublin City University
Dublin, Ireland

Volume 1 of 1

Declaration

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of M.Eng., is entirely my own work and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed: Billy Hogan **ID No.** 92700756

Date: 23/8/96

ABSTRACT

Vocal Tract Acoustic Measurements and their application to Articulatory Modelling

In the field of speech research it is agreed that more real data is required to improve the articulatory modelling of the vocal tract. Acoustic techniques may be used to acquire vocal tract data. The advance of digital signal processing has allowed the development of new experimental techniques that allow fast and efficient measurements.

DSP based measurement systems were set up, and acoustic impedance and transfer function measurements were performed on a wide variety of subjects in DCU's semi-anechoic chamber. The measurement systems are compact and reproducible.

The variation of the wall vibration load was investigated in a wide range of human subjects. The investigation was prompted by the question: Is the wall vibration load important in the study and implementation of vocal tract and articulatory models? The results point to the possible need in acoustic to articulatory inversion, of adapting the reference model to specific subjects by separately estimating the wall impedance load.

Acknowledgements

This thesis would not have been possible without the support, encouragement and interest of my supervisor Dr Ronan Scaife. His wide knowledge of speech research, and his constant efforts to provide the best facilities were a great help. Professor Charles McCorkell and the staff of the school of Electronic Engineering in DCU were always available and helpful.

The "SPEECH MAPS" project introduced me to the European speech research community, and was a privilege to work with such dedicated and enthusiastic people. Thanks especially to Dr Pierre Badin, who gave freely of his valuable time to act as a subject during our experiments in Grenoble. Thanks also to Professor Gunnar Fant, Dr. Shinji Maeda and Mats Bavegard for interesting comments and discussions. Shinji freely supplied the source code for his articulatory model.

I am grateful to Dr Ailbhe Ni Chasaide, and Dr Christer Gobl of Trinity College whose expertise as 'trained' subjects allowed us to resolve the 'twin peaks' question. My employers Ericsson Systems Expertise Ltd, have generously allowed me study leave to complete this thesis.

Thanks to my friends in the speech laboratory Chris, Jeeva and Albert for both professional and non-professional advice. Thanks to the 13th floor crew in DCU, and my friends in Galway for the laughs. Thanks Joe for getting me the PC. Thanks Marie for that alarm call! A special thanks to my family for their love and support.

Finally, this thesis is dedicated to the memory of Sean Collins, friend and fellow post-graduate student. He acted as a subject for these measurements before his tragic death at the age of 24 in June 1995.

List of Symbols

The following is a list of generally used symbols.

c	Velocity of sound in air at body temperature ($= 35000\,\mathrm{cm.s^{-1}}$)
ρ	Density of air ($=0.00114\,\mathrm{g.cm^{-3}}$)
μ	Viscosity coefficient of air ($=0.000186$ dyne. $\mathrm{cm.s^{-2}}$)
f	Frequency (Hz)
ω	Angular frequency ($\mathrm{rad.s^{-1}}$)
s	Complex frequency ($\mathrm{rad.s^{-1}}$), $s = \delta + j\omega$.
L	Inductance ($\mathrm{g.cm^{-4}}$)
C	Capacitance ($\mathrm{g^{-1}cm^{2}}$)
R	Resistance ($\mathrm{g.cm^{-4}s^{-1}}$)
G	Conductance ($\mathrm{g^{-1}cm^{4}s}$)
λ	Coefficient of heat conduction $\left(0.055\times10^{-3}\,\mathrm{cal.cm^{-1}s^{-1}\,deg^{-1}(0^{\circ}C)}\right)$
η	Adiabatic constant (1.4)
c_p	Specific heat of air at constant pressure $\left(0.24\mathrm{cal.g^{-1}\,deg^{-1}(0^{\circ}C, 1atmos)}\right)$
Z_i	Characteristic impedance of the *i*th section
l_i	Length of the *i*th section.
Γ_i	Transfer constant of the *i*th section.
S_i	Circumference of the cross-section of the *i*th section.

Table of Contents

Chapter 1

Introduction

1.1 Introduction

Acoustic and physiological measurements on the human vocal tract apparatus during the production of speech have many applications in speech research. Most importantly, these measurements can be used to assess, evaluate and improve existing vocal tract articulatory models. Despite their usefulness, there is an acknowledged shortage [Fant, 1991] of actual measurement data available to the speech research community. There is a specific lack of measurement data on acoustic loss mechanisms in the vocal tract.

Why such a lack of data? Physiological methods of measuring useful parameters of the vocal tract system have traditionally involved methods using equipment that is bulky, expensive, and inaccessible to the vast majority of speech researchers. So for example we have Fant's famous x-ray measurements [Fant, 1960] of the vocal tract midsagittal plane. These experiments provide useful information, but are dangerous and difficult to conduct. A quote from the descriptions of Drs A.S. Macmillan and G. Kelemen given in [Fant, 1970] serves to illustrate the point.

" The outline of the tongue, the roof of the mouth, and the pharyngeal wall were made visible by giving the subject a mixture of barium and water to which mucilage of acacia was added.....Adhesion to the mucous membranes is very good. This mixture was applied by spatula, and the subject was asked to smooth over the surface with his tongue to get an even distribution." Understandably this sort of data is rare.

Another physiological method used was the x-ray microbeam method [e.g. Guo and Milenkovic, 1993] where x-rays were used to track the motion of small lead pellets placed in strategic positions throughout the subject vocal tract. The equipment used was expensive and the measurements only provide partial data. More modern methods include CAT scans and magnetic resonance imaging [Greenwood et al, 1992] and very good results are being produced; again availability of equipment and exposure times are the main problems.

1.2 Acoustic measurements

If x-ray and scanning methods are too difficult then why can't we obtain vocal tract information by acoustic methods? The simplest acoustic measurement is to record speech directly using a microphone. This sort of measurement can be used to estimate formant frequencies and bandwidths. However it is limited in usefulness because we have no direct knowledge of the glottal source. Various researchers have used excitation-response methods in an attempt to characterise the vocal tract. In these experiments an external source is used to excite the vocal tract, and the response is measured. These experiments include [Fujimura and Lindqvist, 1971; Fant, Branderud and Nord, 1976; and Sondhi 1983]. These researchers achieved good results using different systems, however the procedures had their disadvantages. In the case of the first two mentioned above, the problem was long measurement times requiring the subject to hold a constant VT position for the order of ten seconds. In Sondhi's impedance function measurement the measurement time was low, but the subject had to articulate silently without the benefit of auditory feedback. Acoustic measurements are discussed further in section 2.7.

1.2.1 Acoustic measurements using pseudo-random Digital Signal Processing (DSP) techniques.

With the advent of DSP theory and high speed digital signal processors it has become possible to take a fresh look at acoustic measurements of the vocal tract. Measurement times can be brought down through the use of wideband noise signals as excitation signals and cumbersome analysis of the results can be carried out almost in real-time. Furthermore, the equipment used is widely available so it should be possible for other researchers to reproduce the experiments.

The aim of this thesis is to describe and present the results of DSP based acoustic experiments on the VT. Some of the techniques have been used before, some are new. The results are compared to existing articulatory models.

1.3 Structure of the thesis.

Chapter 2: *Theory and background to speech production modelling and acoustic measurements.* In this chapter the theory of speech production modelling is reviewed. The basic lossless speech production model is described. The transmission line analogy to the vocal tract is described, and electrical-acoustical analogous terms are explained. The modelling of vocal tract energy loss mechanisms by the addition of wall loss components is described. Some previous vocal tract acoustic measurements, their aims and results are reviewed.

Chapter 3: *System Hardware and Software.* This chapter describes the system hardware and software used in the design of the measurement system. A commercially available DSP board plugged into a PC was the hardware platform for the measurements. To provide low noise measurements it was necessary to design and build a complete analog-digital input/output system to interface to the serial port of the DSP board. An overview is given of the software designed to perform, control and analyse vocal tract measurements. The software consists of the user interface and control programs running on a PC, the I/O and DSP algorithms running on a DSP board, and further processing and analysis algorithms running on a PC.

Chapter 4: *Transfer Function Measurements.* In this chapter the theory of transfer function measurements is presented, and the results of measurements of the acoustic transfer function of the vocal tract are shown. The results are compared qualitatively to the output of an existing articulatory model. A new method to determine the closed tract resonance frequency is described.

Chapter 5: *Quasi-Simultaneous Transfer Function and Acoustic Impedance Measurements.* In this chapter acoustic impedance measurements performed at DCU are described. A measurement rig designed and built to allow quasi-simultaneous acoustic impedance/acoustic transfer function measurements is described. Results obtained using the rig are discussed. The impedance measurement device was specially developed at DCU and its operation and calibration are described.

Chapter 6: *The Wall Vibration Load Survey.* A survey of the wall vibration load in 24 human subjects was carried out using the impedance measurement apparatus developed at DCU. The results are discussed and analysed. Some experiments carried

out on a trained speaker are presented, and the results analysed with the help of modelling techniques.

Chapter 7: *Summary and Statement of Conclusions.* In this chapter the VT measurement work carried out at DCU is summarised. Future developments of the measurement techniques are discussed.

1.4 Why articulatory modelling?

Our measurements are designed to supplement, evaluate and improve existing articulatory models, but why use articulatory models in the first place? The main goals of speech engineering research are given below.

1) To synthesise natural sounding speech.
2) To develop low bit-rate speech coders.
3) To develop robust algorithms for speech recognition.
4) The development of other systems such as text-to-speech and medical systems.

In recent years articulatory speech production models have been outperformed by speech systems based on other models e.g. CELP, LPC, formant frequency based models, waveform coders etc. Although a lot of these methods are computationally intensive, the constant improvement of processor capability means that number crunching is becoming less of a problem. However, many researchers believe that the goals above can be achieved by articulatory modelling of the vocal tract system. The reasons for this are explained below

Human speech is produced by passing air from the lungs through the glottis and the vocal tract and/or nasal tract and radiating the sound from the lips and/or nostrils. Precise muscular control of the articulators during this process allows us to combine a wide variety of sound types (e.g. fricatives, plosives, vowels) into continuous speech. Articulatory modelling studies the relationships between the geometry of the vocal tract and its acoustic output. The geometry of the human vocal tract is shaped by the position of the articulators i.e. the jaw, lips, tongue, velum etc. The standard geometrical measure that is used to characterise the VT is the area-function: the cross-sectional area of the VT from the glottis to the lips.

Parametric models of the vocal tract based on a mapping from a mid-sagittal plane as seen from an x-ray, are referred to as articulatory models e.g. [Coker 1967; Lindblom

& Sundberg 1971; Mermelstein 1973; Maeda 1979]. Area functions can be derived from vocal tract shapes parametrised by these models [Maeda 1979]. Alternatively the area function (derived from the measured mid-sagittal view) can be parametrised directly for example, Fant's three parameter model [Fant 1960; Lin 1990]. For all of these parametric models less than 10 parameters are required to describe a vocal tract shape.

Bidirectional transformation between the speech wave and an articulatory model would have the following advantages.

1) The articulatory models require few parameters which interpolate over time. The vocal tract shape is also relatively slowly time-varying, so for continuous speech there is a low frame rate (~100Hz [Schroeter and Sondhi, 1989]) . This would lead to low bit-rate speech coders.
2) It has been shown that good quality speech can be produced through phoneme level control of an articulatory model [Flanagan, Ishizaka and Shipley, 1975].
3) Articulatory models, because of their direct relation to the way real humans speak, have medical applications. For example, A view of vocal tract shapes during speech production could be useful in speech therapy for people with hearing disorders.

The advantages are summarised in the diagram of [Jeevakumar, 1993] given below.

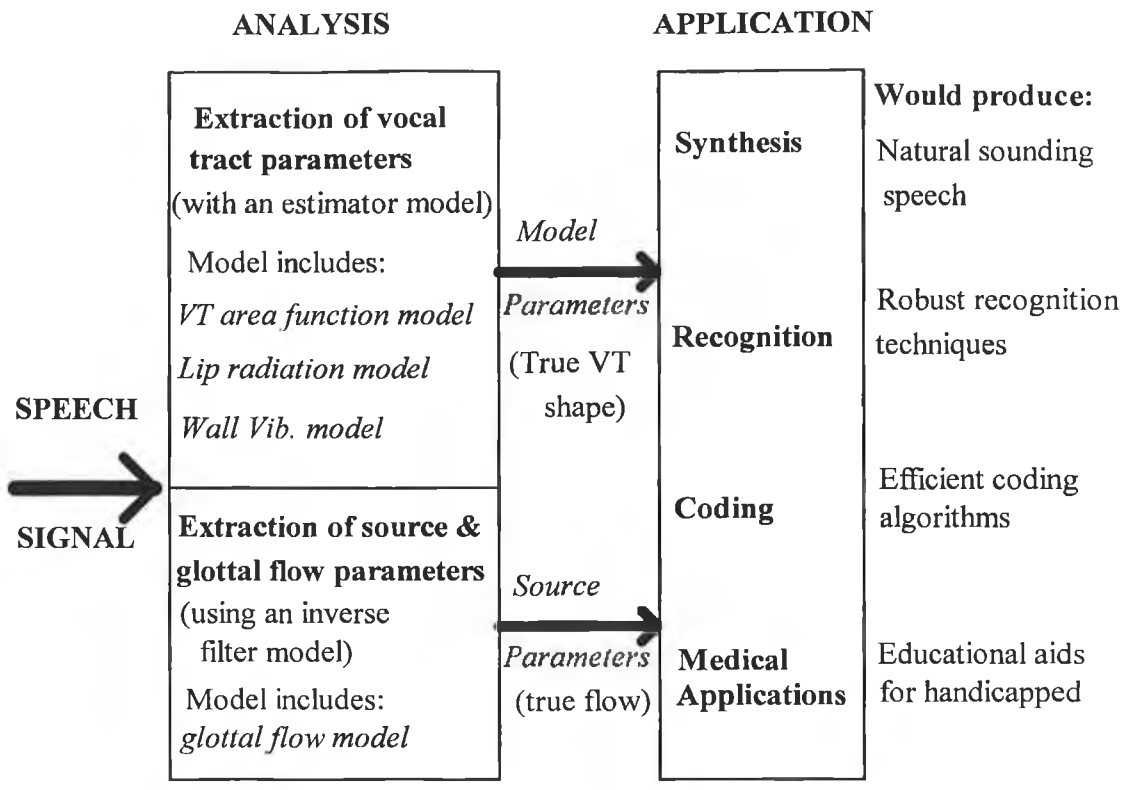


Figure 1.1: Diagram showing the applications and benefits of using the physiological data of the human speech production system in speech analysis.

If articulatory systems have so many advantages why isn't a greater proportion of speech engineering research effort concentrated in this area? The great problem with articulatory modelling is the inverse problem. There is a many-to-one mapping between the vocal tract geometry and its acoustic output. In simple terms, different vocal tract shapes can produce the same sound. So while it is not difficult to produce a speech sound given an area function, it is very difficult to produce the 'correct' area function given a speech sound. How can this problem be tackled? If we take a speech wave produced by a given utterance, and examine the possible set of area functions that could have produced the same utterance we will see that many of the area functions can be eliminated because they represent positions of the articulators that are impossible or extremely difficult to achieve. There are also kinematic constraints whereby transition from one area function to another would not be possible within the given time. So constraints must be added to minimise the possible set of area functions. These are the problems being dealt with by the European Community ESPRIT basic research project entitled SPEECH MAPS.

1.5 The SPEECH MAPS project.

Most of the work included in this thesis was part of DCU's contribution to the SPEECH MAPS project. The goal of the project is to construct a speaking robot, the *Articulotron*, that is capable of learning how to speak from acoustic samples of speech.

The project, involving universities and research institutions from all over Europe is divided into four working parties.

WP1: From speech signal to acoustic sources.

WP2: *From speech signal to vocal tract geometry.*

WP3: Dynamic Constraints and motor control

WP4: Perceptual processing for gesture recovery

The work conducted in DCU was part of WP2. The goal of WP2 is to provide WP3 with supporting data, forward models and background knowledge useful for constructing the articulotron.

The specific aims of WP2 were:

1) The acquisition of vocal tract data by geometric and acoustic measurements.
2) The improvement of the mid-sagittal dimension to area function conversion.
3) The improvement of the plant, i.e. an articulatory model associated with a codebook specifying articulatory-acoustic relationships.

The work presented in this thesis was concerned mainly with part one, and results were fed into part three. In the next chapter a speech production model is discussed, and some previous acoustic measurements of the vocal tract are reviewed.

Chapter 2

Theory and background to Articulatory Modelling and Acoustic Measurements

2.1 Introduction

In speech engineering, the principal goal of acoustic measurements of the vocal tract is to supplement and improve the mathematical modelling of the vocal tract. Realistic acoustic-to-articulatory and articulatory-to-acoustic mappings can only be obtained through an accurate vocal tract model. In this chapter the theory of speech production and articulatory modelling is reviewed.

Acoustic models of the vocal processes may be developed in either the time domain or the frequency domain. Time domain models involve the solution of the wave equations representing the flow of air through the vocal tract. In frequency domain modelling, the vocal apparatus is represented as an acoustic source connected to a variable acoustic filter. According to Maeda [Maeda, 1982] "the acoustical elements, such as radiation impedance at the mouth opening and boundary losses are specified more accurately in the frequency domain than the time domain. The dynamic events, such as voice excitation on the contrary are described more effectively in the time domain". The basis for frequency domain modelling of the vocal tract has been developed since the 50's, for example [Stevens, Kasowski and Fant 1953]; [Fant 1960], [Flanagan 1972]. In recent years, efforts to improve the accuracy of such models has included work on such subtle and difficult areas as source-tract interaction, nasalisation, and vocal tract energy loss mechanisms [Speech Maps report (ESPRIT/BR no. 6975), 1994]. Measurements are essential for progress in the modelling of these effects.

2.2 The Speech production system.

The human vocal tract consists of the lungs, tracheal tubes, larynx, pharyngeal cavities, oral chambers and nasal passages, as shown in figure 2.1. The vocal tract is

the principal path for sound transmission. It begins at the glottis and terminates at the lips. The nasal tract is a side branch that begins at the velum and terminates at the nostrils. During the production of nasalised sounds the nasal tract is coupled to the vocal tract by lowering the velum.

In the basic source-filter model for voiced sounds, the source is the modulated airflow passing through the glottis as the vocal folds are vibrating in an almost periodic way. Since this is the source for voiced sounds, it is termed the *voice source*. The vocal tract acts as a variable acoustic filter. Variation of the vocal tract shape changes the spectral response (also known as the formant pattern) of the filter, and so different sounds are produced and radiated at the lips and/or nostrils.

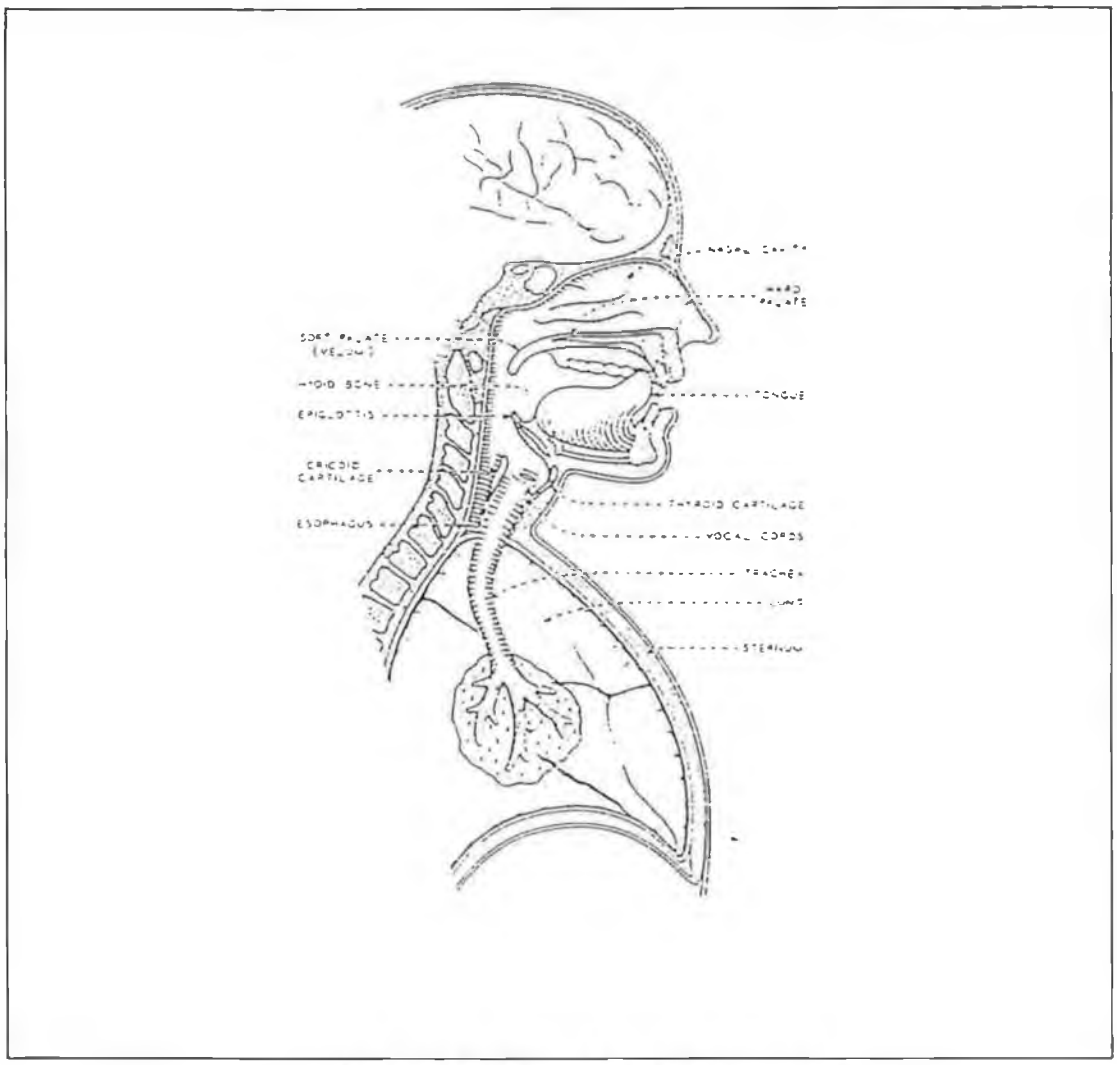


Fig 2.1 The human vocal apparatus

(Diagram from [Flanagan, 1972])

Quoting [Fant 1960]. "The frequency selective characteristics are introduced by the process of multiplying the amplitude of each harmonic |S(f)| of the source spectrum by the value of the appropriate gain factor |T(f)| of the filter function at the frequency f."

$$|P(f)| = |S(f)||T(f)| \tag{2.1}$$

Of course the vocal tract may be excited by sources other than the voice source. For voiceless fricatives (e.g., /s/), the source is at the point of the constriction in the vocal tract and consists of turbulent airflow. For voiceless plosives (e.g., /p/), the source is at the point of closure and consists of a rapid release of the air pressure built up behind the total constriction.

The basic source-filter assumption that the source of excitation and vocal tract are independent allows us to discuss the transfer function of the vocal tract as the vocal tract *filter* regardless of its excitation by any of the possible sources. The length of the vocal tract is about 17.5cm for an adult male, and 14cm for an adult female. The representation of wave motion through the vocal tract is generally simplified by assuming one-dimensional wave propagation. This assumption is considered valid up to frequencies of around 4-5kHz, when the cross-dimensions of the vocal tract become comparable to the wavelength of the acoustic signal.

2.2.1 The acoustic wave equations and a digital speech model

The equations for the propagation of acoustic energy through the vocal tract are derived from the fundamental laws of physics. In particular, the conservation of energy, the conservation of momentum, and the conservation of mass along with the laws of thermodynamics and fluid mechanics [Rabiner & Schafer, 1978]. One form of the wave equation was derived for a lossless tube by [Portnoff, 1973].

$$-\frac{\partial p}{\partial x} = \rho \frac{\partial \left(u/A \right)}{\partial t} \tag{2.2a}$$

$$-\frac{\partial u}{\partial x} = \frac{1}{\rho c^2} \frac{\partial (pA)}{\partial t} + \frac{\partial A}{\partial t} \tag{2.2b}$$

Where:

p=p(x,t) : The sound pressure in the tube at position x and time t.

u=u(x,t) : The volume velocity of air in the tube at position x and time t.

ρ : The density of air in the tube.

c : The velocity of sound in air.

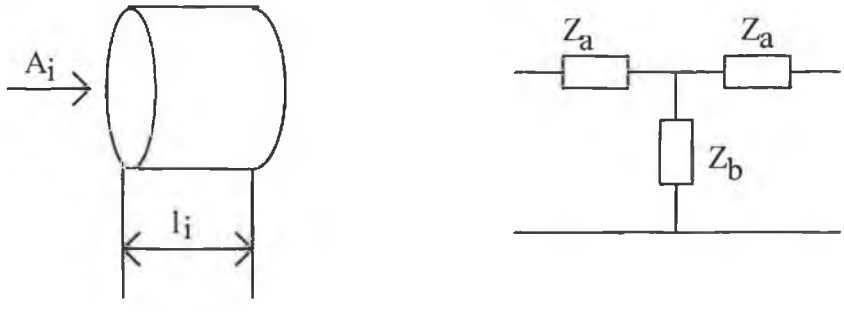
A=A(x,t) : The cross sectional area of the tube at position x and time t.

The solutions to these equations would be difficult for a real vocal tract. Complete knowledge of the area function is required as well as boundary conditions at the lips and glottis. The equations become more complicated when loss factors are taken into account. The solutions to equations 2.2a and 2.2b above for a small tube of constant cross-sectional area A are given in appendix A.

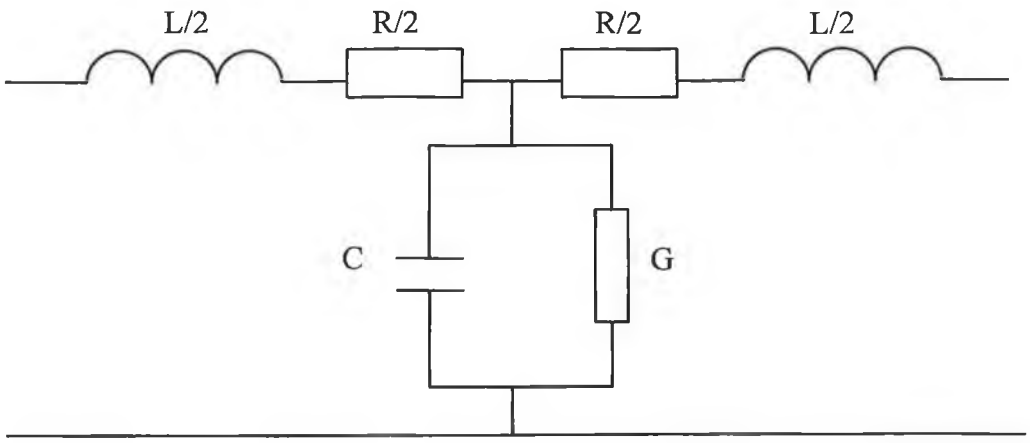
2.3 The Transmission line analog of the vocal tract

The cylindrical tube sections described above can also be represented using equivalent electrical circuits as shown in fig 2.2 below. This configuration is particularly useful for description of frequency domain losses in the model. The most common configuration used is the T-network. The T-network elements may be derived given the dimensions of a tube section using acoustic transmission line theory [Flanagan 1972].

The equivalent circuit elements are the per-unit-length inductance, capacitance, series resistance and shunt conductance: L, C, R, and G respectively.



a b



c)

Fig 2.2 a) A hard-walled cylindrical section of length l_i and area A_i

b) Equivalent T-network of a)

c) Lumped version of b), valid for short l_i.

The components (Z_a and Z_b) of the T-network above have been shown to be [Flanagan 1972]

$$Z_a = Z_o \tanh\left(\Gamma_i / 2\right) \tag{2.3}$$

$$Z_b = Z_o / \sinh(\Gamma_i) \tag{2.4}$$

Where the transfer constant Γ_i and the characteristic impedance Z_o of a section are related to the distributed parameters RLGC per-unit-length and to the section length l_i by the classical expressions:

$$\Gamma_i = \sqrt{(R + j\omega L)(G + j\omega C)}.l_i = \gamma.l_i = (\alpha + j\beta).l_i \tag{2.5}$$

$$Z_o = \sqrt{\frac{R + j\omega L}{G + j\omega C}} \tag{2.6}$$

Where γ is the complex propagation constant consisting of α the attenuation constant and β the phase constant. For small losses ($R << \omega L, G << \omega C$) the attenuation and phase constants are approximately:

$$\alpha \cong \frac{R}{2}\sqrt{\frac{C}{L}} + \frac{G}{2}\sqrt{\frac{L}{C}} = \alpha_R + \alpha_G \tag{2.7}$$

$$\beta \cong \omega\sqrt{\frac{L}{C}} \tag{2.8}$$

We see that the attenuation constant α is split into two terms α_R representing the attenuation due to frictional losses, and α_G representing the attenuation due to heat conduction losses.

2.4 Acoustic-electrical analogous terms

Throughout this thesis acoustic-electrical analogies are used. Here a brief description of the meaning of these analogies is given. If sound pressure, P, is considered analogous to voltage, and acoustic volume velocity, U, analogous to current then the electrical circuit elements R, C, L, G all take on meaning in the acoustic domain as explained below. Derivations of the equations given below from first principles may be found in [Flanagan 1972].

2.4.1 The acoustic 'R'

The acoustic 'R' represents a power loss proportional to U^2 and is the power dissipated in viscous friction at the tube walls for a uniform tube. The average power dissipated per-unit-length is given by:

$$P.S = \frac{1}{2}U_m^2 R_a \tag{2.9}$$

where:

$$R_a = \frac{S}{A^2}\sqrt{\frac{\omega\rho\mu}{2}} \tag{2.10}$$

R_a is the per-unit-length acoustic resistance, S is the tube circumference, μ is the viscosity coefficient, and ρ is the air density.

2.4.2: The acoustic 'L'

The electrical inductance is analogous to the mass of air in the tube. The acoustic inertance per-unit-length is given by:

$$L_a = \frac{\rho}{A} \tag{2.11}$$

2.4.3 The acoustic 'G'

The acoustic conductance provides a power loss proportional to the square of the local sound pressure. This loss is caused by heat conduction at the walls of the tube. The equivalent conduction per-unit-length of the uniform tube is given by:

$$G = S\frac{\eta - 1}{\rho\, c^2}\sqrt{\frac{\lambda\omega}{2c_p\rho}} \tag{2.12}$$

where λ is the coefficient of heat conduction, η is the adiabatic constant, and c_p is the specific heat of air at constant pressure.

2.4.4 The acoustic 'C'

The acoustic capacitance or compliance represents the compressibility of the volume of air contained in the uniform tube. The equivalent capacitance per-unit-length of the uniform tube is given by;

$$C = \frac{A}{\rho\, c^2} \tag{2.13}$$

Using the equations for the circuit elements (2.10-2.13) in the representations of α_R and α_G given in equation 2.7, we get:

$$\alpha_R = \frac{1}{2\rho\, c}\frac{S_i}{A_i}\sqrt{\frac{\mu\omega\rho}{2}} \tag{2.14}$$

$$\alpha_G = \frac{S_i(\eta - 1)}{2cA_i}\sqrt{\frac{\lambda\omega}{2c_p\rho}} \tag{2.15}$$

Thus we can derive the ratio between the heat conduction loss and the frictional loss.

$$\frac{\alpha_G}{\alpha_R} = (\eta - 1)\sqrt{\frac{\lambda}{c_p\mu}} = 0.444 \tag{2.16}$$

for the values of the constants given at the start of this thesis. So the conduction losses are approximately half the frictional losses, and this ratio is independent of the vocal tract geometry.

2.5 Combining the T-network representations into a speech model.

The basic T-networks shown above are combined just as in the digital model to give a complete model of the vocal tract system. It has been shown [Fant 1960], that the pressure and volume velocity at the input of the i^{th} t-network is related to the pressure and flow at the input of the $(i-1)^{th}$ t-network by the transmission matrix as follows:

$$\begin{bmatrix} P_i \\ U_i \end{bmatrix} = \begin{bmatrix} \cosh(T_i) & Z_0 \sinh(T_i) \\ \sinh(T_i)/Z_0 & \cosh(T_i) \end{bmatrix} \begin{bmatrix} P_{i-1} \\ U_{i-1} \end{bmatrix} \quad (2.17)$$

[Lin, 1990] has developed fast numerical methods to compute transfer functions using a transmission line model.

2.6 Wall losses

The energy losses that occur because of heat conduction and viscous friction are much smaller than the losses due to vibration of the non-rigid vocal tract walls [Fant 1972]. How do we take account of the vibration losses in the transmission line analog? In [Ishizaka et al, 1975] an extra shunting branch is added to the basic T-network. In classical mechanical terms this is equivalent to representing the vibrating walls by localised spring-mass-damper systems. The effect of the wall vibration load is to cause an upward shift in the formant frequencies especially the low frequency F1, and an increase in formant bandwidths. The shunting branch is usually defined as a series RLC circuit, with an impedance of:

$$Z_w = R_w + j\left(\omega L_w - \frac{1}{\omega C_w}\right) \quad (2.18)$$

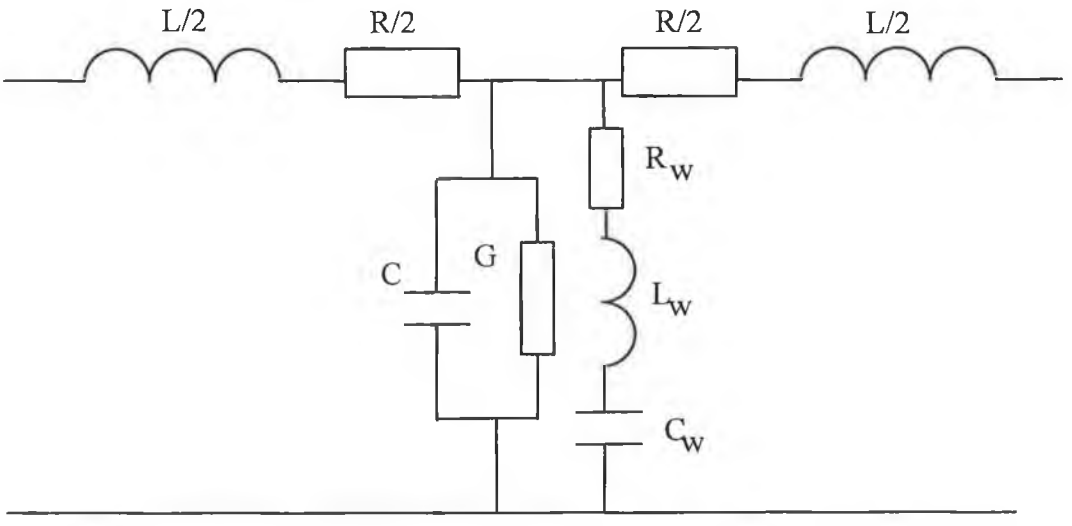


Figure 2.3: The basic T-network with wall loss shunt elements added

Here R_w is the wall resistance and represents heat conduction and friction losses in the vocal tract walls. L_w is the wall inductance and represents the mass or inertance of the vocal tract wall. C_w is the wall capacitance, and represents the compliance or

elasticity of the walls. The wall elements can either be included in a distributed element representation [Ishizaka et al, 1975], or inserted as lumped element shunts as in [Fant et al, 1976]. Fant has shown that most of the wall losses in the vocal tract occur in two locations, the first and foremost in the vicinity of the pharynx, and the second near the lips. This result makes the concept of two lumped element shunts, one at each location, a viable proposition. In chapter six we present measurements of these lumped wall elements.

2.7: Some previous vocal tract acoustic measurements

Input impedance measurements of the vocal tract system are not new. As early as 1958 Van den Berg [Van den Berg, 1958], was conducting measurements on the subglottal input impedance of the cadavers of large dogs and humans. This area was revisited again in 1976 [Ishizaka, Matsudaira and Kaneko, 1976]. The purpose of these measurements was to facilitate the development of a model of the subglottal system which could be used in an overall vocal tract system model. The results of the two sets of experiments differed considerably, the main reason being that the later experiments were conducted on live semi-laryngectomised subjects. In this operation the top of the trachea is stiched closed, and an opening for breathing is made on the neck about 2cm below the closed end.

In the measurements of Ishizaka et al an excitation adaptor was attached to this opening, and the pressure signal response to a frequency sweep was measured via two microphones placed either side of a known acoustic impedance. Standard circuit theory was then used to solve for the unknown subglottal input impedance. The results showed subglottal impedance peaks at frequencies of 640Hz, 1400Hz and 2100Hz. A major drawback of the measurement system was that the adaptor actually blocked the subjects breathing apparatus for 30 seconds at a time, so the subjects had to maintain inflated lungs throughout the measurement. Three measurements of 30 seconds each were required to cover a frequency range of 3kHz, so consistency of measurement becomes a problem.

It is easy to see how this measurement system could be improved using random noise sequences as excitation signals. Measurement times for a frequency range of over 4kHz could be brought down to less than a second. The results would become more

consistent, and it would be possible to obtain measurements with varying degrees of lung inflation.

2.7.1 Impulse response measurements of the vocal tract

Vocal tract lip impulse response measurements were first attempted by M.R. Schroeder at Bell laboratories, [Schroeder, 1966]. The method was later developed by Sondhi, [Sondhi, 1983]. Schroeder showed that if the area function of the vocal tract was represented by a Fourier series, and small perturbations were applied to the area function, the resonance frequencies of the vocal tract were also perturbed in a corresponding manner. For example, in the case of a tube of uniform cross sectional area A_0, closed at one end and open at the other, the perturbation $\delta A(x)$ could be represented by a cosine Fourier series:

$$\delta A(x) = A_0 \sum_{m=1}^{\infty} a_m \cos(\pi m x / L) \qquad (2.19)$$

where L is the length of the tube. He found that each *odd* coefficient (a_m, $m = 1, 3, 5\cdots, \infty$) in the Fourier series expansion of the area perturbation, affected one and only one of the resonance frequencies, the even Fourier coefficients of equation 2.19 have no first order effect. These relationships suggest that the set of area functions can be restricted to assure a unique relationship with the resonance frequencies, at least for small perturbations. If only antisymmetric perturbations about the midpoint of the tract are allowed, then all even terms in the Fourier series expansion of equation 2.19 are identically zero and the odd coefficients can be determined from the frequency shifts from the uniform tract.

Schroeder was able to show that for a uniform tube that is either closed or open at both ends, resonance frequency shifts are directly related to the even Fourier coefficients of the area function perturbation. He then reasoned that if it was possible to measure transfer functions for both open and closed lip conditions, that it would be possible to obtain both the odd coefficient and the even coefficient parts of the Fourier series, and thus the area function could be reconstructed.

Although it was not practical to perform simultaneous measurements of that type at the time, it had been shown by Mermelstein [Mermelstein, 1966], that the impedance function of the vocal tract measured at the lips has a unique relation to the resonance

frequencies of the tract for both open and closed lips. The zeros of the impedance function are equal to the open-lip resonance frequencies of the vocal tract, and the poles of the impedance function are equal to the closed-lip resonances of the vocal tract. It should be noted here that the limitations of this theory was that it was only applicable for *small* perturbations of a *lossless* vocal tract.

Schroeder performed some measurements on the lip-impedance of human beings. This was done by sending a pulse of duration 10ms and bandwidth 4kHz down a long tube (about 2m), closed at one end (the driving end), and open at the other. The subject articulated silently into the tube. The incident and reflected waves were measured using two condenser measurement microphones inserted into the tube. The Fourier coefficients of the input impedance at the lips were given by the formula:

$$Z_n = Z_0 \frac{F_{1n}\mathrm{Sin}(2\pi n f_0 \tau_1) - F_{2n}\mathrm{Sin}(2\pi f_0 \tau_2)}{F_{1n}\mathrm{Cos}(2\pi n f_0 \tau_1) - F_{2n}\mathrm{Cos}(2\pi f_0 \tau_2)} \qquad (2.20)$$

where τ_1 and τ_2 are the travel times for sound from microphones 1 and 2 respectively, to the lips. f_0 is the fundamental pulse rate (100Hz), and Z_0 is the characteristic impedance of the measurement tube, and F_{1n} and F_{2n} are the complex Fourier coefficients of the two microphone signals.

Area functions were then reconstructed from the measurement impedance functions. The experiments showed that reasonable matches could be obtained from measurements on hard walled objects with known area functions. The measurements on people showed a greater deviation from area functions measured using x-ray techniques. Schoeder summarised the possible limitations of the method as follows:

"*Can a seal be found that is tight enough and allows unrestrained articulation?...If a fixed (vocal tract) length is assumed, a variable length may lead to errors in the derived area function, especially near the glottis.....is the 10ms sampling of the impedance function sufficient for rapid movement of articulators during phoneme transitions....more significantly are there any significant effects of the shape of the cross-sectional area on the singularities of the impedance function? Are losses important? These questions imply a possible inadequacy of Webster's horn equation.*"

He also states *"Perhaps the most important question concerns whether impedance measurements at the lips can give useful unique area functions for finite nasal coupling or whether auxiliary (but simultaneous) measurements are required to determine the nasal coupling parameter."* These problems are listed here, because even though our measurements of lip impedance used a different method, similar type problems arose.

2.8 Summary

In this chapter the theory and background to vocal tract modelling and acoustic measurements were discussed. The transmission line model of a lossy vocal tract, and acoustic-electrical analogous terms were described. A summary was given of some previous acoustic measurements that used concepts similar to those of the measurements performed in this thesis. In the next chapter the hardware and software setup of the measurement systems is described in greater detail.

Chapter 3

System Hardware and Software

3.1 Introduction

The purpose of this project was to collect and process human articulatory data. To this end measurement systems had to be set up with the speed to handle numerically intensive operations and the flexibility to deal with such requirements as simultaneous input/output of a number of data channels, variable sampling rate, and user feedback.

The hardware for the project was based on a data-acquisition system with digital signal processing capabilities. This allowed the measurements to be controlled from a PC, and numerically intensive computations to be handled quickly. The speed of computation determines the amount of feedback a subject can derive from the measurement system. Waiting five or six seconds after a measurement has been made for the results to be displayed is acceptable whereas waiting forty or fifty seconds is not. In this chapter an overview of the hardware and software designed to perform, control, and analyse vocal tract measurements is given. The software for this type of measurement system can be divided into three distinct parts.

- The user interface and control programs running on a PC
- The I/O and DSP algorithms running on the LSIC30 board
- Further processing and analysis algorithms running on a PC.

This chapter concentrates on the first two parts. The further processing and analysis algorithms are discussed in Chapters 4 and 5.

Our needs for the measurement system were specific. The measurements had to follow particular protocols with control often switching between the PC and the DSP board. The DSP algorithms running on the LSIC30 board had to be implemented as efficiently as possible, so that results could be quickly transferred to the PC and

displayed. The whole process from initiation of a measurement to display of the final results had to be automatic, with the user controlling timing by keyboard hits.

3.2 Choosing the hardware setup:

Two commercial DSP/data-acquisition cards were available in DCU speech lab.

- The IBM Audio Capture and Playback Adapter (ACPA)
- The Loughborough Sound Images TMS320C30 (LSIC30) System Board.

Both of these cards were investigated to decide which one would be the core of the measurement system. The IBM ACPA is a DSP/data-acquisition card specifically designed for audio applications. It is based on the popular TMS320C25 digital signal processor, and its main features are:

- Dual on-board 16-bit A/D and D/A converters.
- 16K of on-board RAM

The ACPA plugs into an expansion slot in the back of a PC, and it controlled from the PC via a memory mapped register in the PC's I/O space. A disadvantage of the ACPA is that its applications program interface (API) is not well developed. An applications programmer who wished to use the ACPA for more than very simple applications, would have to write a library of interface functions to communicate with the board before the application could be implemented. The interface library would have to include functions to initialise the board, reset the board, transfer data in various formats to and from the board, and functions to hold and halt the board. The development of an API for the ACPA is a significant job in itself. The LSIC30 board has an advanced API, and for this reason was chosen above the ACPA as the applications development platform for our system. The LSIC30 board is based on the TMS320C30 digital signal processor, and a brief overview of this chip is now given.

3.3 Overview of the TMS320C30

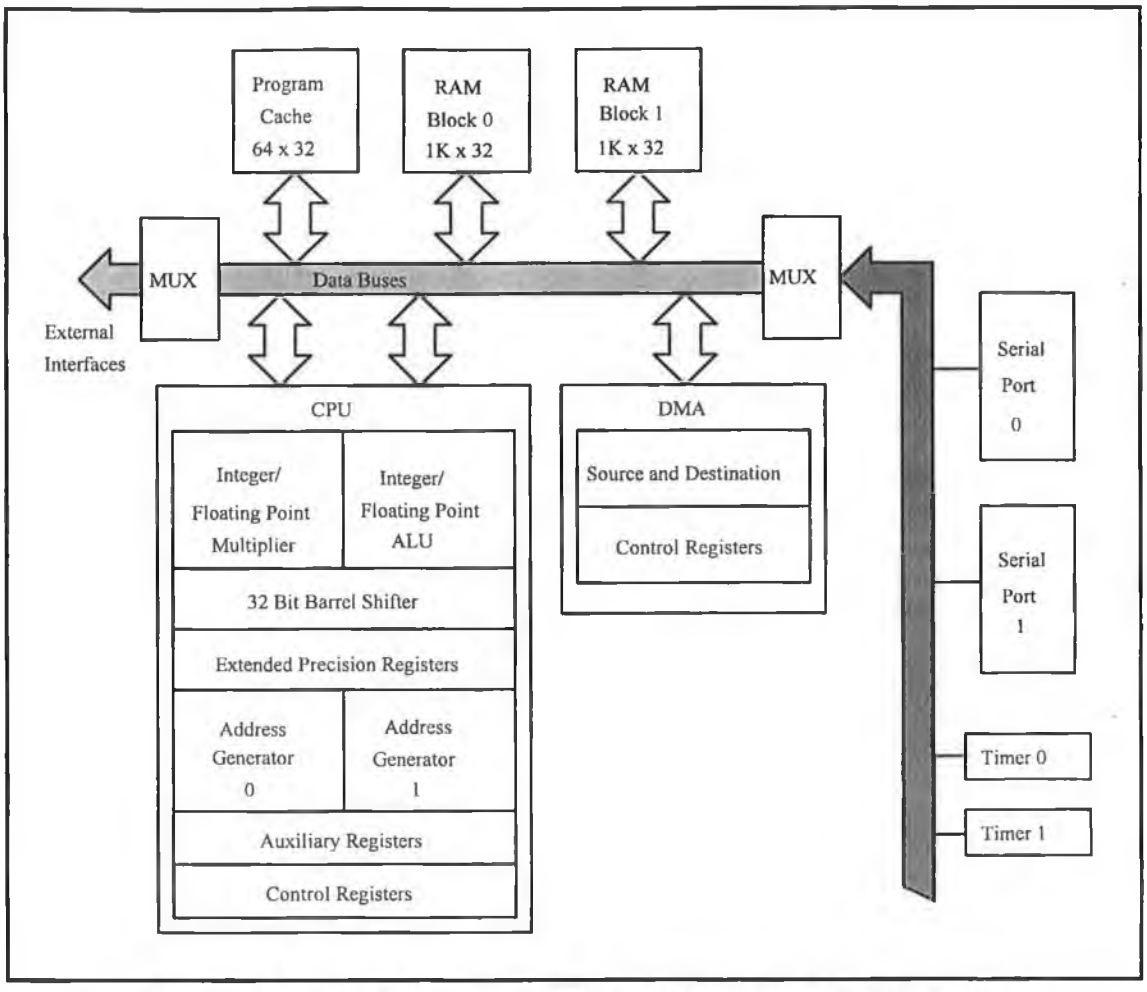


Figure 3.1: Block Diagram of the TMS320C30 Architecture

The TMS320C30 is a 32-bit floating point processor. It is the third generation of the TMS320 family and is one of the most popular programmable DSPs currently available on the market. Some of its most important features are:

- 60ns single instruction cycle time
- 33.3 MFLOPS (million floating-point operations per second).
- 32-bit instruction and data words, 24-bit addresses.
- Two 1K by 32-bit single cycle dual access on-chip RAM blocks.
- 64 by 32-bit word instruction cache.
- Two serial ports supporting 32-bit transfers.
- Separate program and data buses.

3.3.1 Memory organisation

The total memory space of the 'C30 is 16 million 32-bit words. Programs, data, and I/O space are contained within this address space. The 2K of on chip RAM is capable of supporting two CPU accesses in a single cycle. A 64 by 32-bit instruction cache is provided to store often repeated sections of code, and greatly reduces the number of off-chip accesses necessary.

3.3.2 Peripherals

The 'C30 peripherals include two timers and two serial ports. All the peripherals are controlled through memory mapped registers on a dedicated peripheral bus.

3.4 The Loughborough Sound Images TMS320C30 system board.

The LSIC30 system board is a digital signal processing card based on the Texas Instruments TMS320C30 ('C30) digital signal processor. It plugs into a full length expansion slot on a PC. Its main features are:

- The TMS320C30 floating point digital signal processor
- Dual on-board 16-bit A/D and D/A converters
- 128K on-board RAM
- A PC interface

The system board comes with extensive software from LSI. This includes:

- The SPOX API.
- The TI assembler/linker and the TI C-compiler.
- A C-language interface library.
- Two debug monitors.

3.4.1 System board memory

The on-board memory is divided into four banks accessible through the memory address space of the 'C30. Bank 0 contains 64K words of 0 wait-state memory. Bank 3 contains 64K words of 1 wait-state memory. The other 2 banks of 64K words one wait-state memory are not populated on the standard board. The bank 3 memory is dual ported and is used to transfer data between the 'C30 and the PC without halting the 'C30. Data may also be transferred between the PC and Bank 0 but the 'C30 processor must be halted during the transfer.

3.5 The serial-port interface

One of the main factors to be taken into account in improvement of the noise performance of the system was the conversion accuracy of the analog input/output system. To quote the system board users guide "Because of system noise and device non-linearity, effective performance is less than 16 bits. Actual performance is dependent on the PC being used". Our own measurements on LSI boards in a number of PCs showed that the best performance that could be obtained was 10 bits; 6 bits were being lost due to system noise. The main cause of noise is crosstalk and interference with other components on the circuit board (including the 'C30 chip itself). In general, with high resolution A/Ds and D/As system design problems such as ground path resistance and power supply quality become very important. It was felt that to improve the bit resolution, it was necessary to design an alternative analog interface that was isolated from the rest of the system as much as possible.

Fortunately, as well as an analog interface connector, the 'C30 board also has a 'serial expansion connector' on the endplate. This provides a fully buffered version of the serial-ports of the TMS320C30. The two channels of the serial-port can be used to transfer data to and from the TMS320C30. It was decided to design and build an external analog interface that would connect to the serial-port of the TMS320C30 and thus bypass the internal analog interface altogether. This would, it was hoped, increase the bit resolution of the system. There was also the added advantage that two channels could now be played back or recorded simultaneously. The implementation of this idea involved a number of steps.

1. Design of the circuit
2. Design of the software to program the serial-port to interface to the circuit
3. Breadboard construction and testing of the circuit
4. Upon completion of testing, design of a PCB layout for the circuit using CAD tools.
5. Manufacture and testing of the PCB
6. Designing and construction of a housing case for the PCB, with connectors for all external signals and power supplies.

All of these stages are described in detail in appendix B (B1.4-B1.6). As the operation of the circuit depends so much on the operation of the serial-ports of the TMS320C30, sections B1.1-B1.3 in appendix B also provide a brief description of the serial ports and their configuration in the system.

3.6 Choosing the system software setup.

The following software tools are available with the LSIC30 board:

- The SPOX API.
- The TI assembler/linker and the TI C-compiler.
- A C-language interface library.
- Two debug monitors.

The choice of software tools means that there are three distinct ways to write DSP applications using the LSIC30 board. They are:

- Assembly based
- Plain C based
- SPOX based

These approaches were investigated in order to decide which would be the most suitable for the measurement system. A brief description of each is given in appendix B (section B1.7).

It was decided to use the assembly based approach, as this was the most efficient and flexible method of implementing the algorithms. MATLAB was used to display the results, the board control programs were written in C, and the programs to run on the board were written in TMS320C30 assembly language. Programming the TMS320C30 in assembly language is described in appendix B (Section B2.2).

3.7 Assembly programs for dual transfer function/lip impedance measurements

An assembly program was written in Texas Instruments assembly language, to control the operation of the simultaneous record and playback for both the transfer function and lip-impedance measurements. It communicates with the Burr Brown A/D and D/A converters via the serial port for the transfer function measurement. For the impedance measurements it uses both channels of the onboard converters (see chapters 4 and 5).

3.7.1 Operation of the on-board measurement program

1) Initialises the serial ports of the 'c30 board.

2) Configures the serial port timers to provide the bit clock and the sample rate clock for the external analog interface.

3) Operates the simultaneous record and playback using interrupt service routines. The first interrupt service routine controls the input/output for the transfer function measurement. The interrupt service routine runs on the transmit interrupt of the serial port. This interrupt is controlled by the sample rate clock, so one interrupt occurs every sampling period. The interrupt service routine reads a 32 bit word out to the data transmit register of the serial port, and reads a 32 bit word in from the data receive register of the serial port. The 32 bit word consists of channel A in the top 16 bits and channel B data in the bottom 16 bits. The second interrupt service routine controls the input/output for the impedance measurement. One output and two input channels are required for the impedance measurements. The onboard A/D and D/A converters are used. As before both of the 16 bit input channels are combined into one 32 bit word with channel A in the top 16 bits and channel B in the bottom. This makes it more efficient to read the data up to the PC, where the channel separation takes place.

4) When the simultaneous playback and record are finished the program performs a circular cross correlation of the recorded sequence and the playback sequence for the transfer function measurement and stores the result in on-board memory. For sequences of length 2047 samples, over 12 million calculations are required. This can be done on the 'C30 card in approximately 3 seconds.

5) When the calculations are finished the program informs the C program running on the PC, and the C program then reads up all the results from the on-board memory to PC memory.

3.7.2 Continuous noise playback

A program was written in assembly language to provide a continuous noise playback. This is required in the experiment so that the subject can listen to the response to the excitation of the VT though headphones and thus find the point on the neck that gives the loudest response (see Chapter 4). The setup is similar to the measurement program described in section 3.7.1 except that the interrupt service routine loops continuously. Termination of the program is controlled by a keyboard hit from the user. Control then reverts to the controlling program running on the PC.

3.8 The PC based measurement controlling program

The program controlling the dual transfer function/ lip impedance measurements is written in C and has the the following functions

1) Initialise the 'C30 board.
2) Load the assembled and linked COFF format continuous noise program (described in section 3.7.2) down to the on-board memory of the 'C30 board, and start its execution.
3) Load down the pseudo-random noise sequence to the onboard memory of the 'C30 card.
4) Control the termination of the continuous-noise program when required by the user. Load down the on-board measurement program (described in section 3.7.1) and start its execution.
5) Wait until the on-board measurement program has finished executing and read the results up to PC memory.
6) Convert the results to matlab format, and to write them to datafiles (with the extension .mat).

3.9 The Matlab interface script.

The whole measurement procedure, from initiating a measurement to display of the results on the screen is done using Matlab. The script file written to control the process for the dual transfer function/lip impedance measurements has the following functions.

1) Call the executable measurement control program described in section 4.5 and supply it with arguments to generate the results datafiles
2) When the measurement has been taken read the results into the Matlab workspace.
3) Analyse the data, e.g. using Fourier Transforms to tranform the results to the frequency domain
4) Plot the transfer function and the impedance function on the screen.

3.10 Generation and properties of the excitation sequences.

The excitation sequences used in these experiments were pseudo-random shift register sequences, also known as maximal length sequences or Galois sequences. Pseudo-random sequences have properties that made them more useful for our purposes than other standard excitation sequences such as sine-sweep or white noise. These sequences have been studied extensively because they find use in communications systems, encipherment, and error-correcting coding. The theory of generation of pseudo-random sequences is discussed briefly in appendix B 1.7. A C-language simulation of feedback shift registers was written to generate all of the sequences used for the measurements in this project.

3.11 Summary

This Chapter has discussed the hardware and software aspects of the system used for the measurements presented later in this thesis. The core of the system is the Loughborough Sound Images TMS320C30 system board. In the next three chapters vocal tract acoustic measurements using the systems described here are described.

Chapter 4

Transfer Function Measurements

4.1 Introduction

Chapters 1 and 2 discussed the reasons for acoustic measurements, and showed how they can add to the study of articulatory modelling. Chapter 3 presented the hardware and software setup of the measurement system. In this chapter the theory of transfer function measurements is presented, and the results of measurements of the acoustic transfer function of the vocal tract are shown. The measurements are compared qualitatively to the output of the Maeda articulatory model [Maeda, 1982] for five vowels. Similarity in the overall spectral shape is observed.

The acoustic parameters of the speech signal that are considered most useful for articulatory studies are the frequencies and bandwidths of the vocal tract resonances. These parameters can be estimated directly from the frequency domain representation of the speech wave, although in this case the excitation is unknown and can only be estimated. For voiced sounds, the excitation is caused by the modulation of airflow from the lungs by the quasi-periodically vibrating vocal folds. This excitation introduces a harmonic structure into the output spectrum which tends to smear the formant bandwidths, and make formant frequency estimation more difficult. The method described below avoids this problem, by injecting the excitation source externally at a point a few cm above the level of the glottis. With the input and output both known, the VT transfer function can be calculated with greater accuracy.

The method used is the fast acoustic transfer fuction measurement procedure of Djeradi et al [Djeradi et al, 1991]. The vocal tract (VT) is externally excited at the level of the thyroid cartilage using a pseudorandom binary sequence outputted to an electromagnetic shaker. Simultaneously the response is measured at the lips using a condenser microphone. The input and output sequences are stored on a PC fitted with a signal processing card, and are used to calculate the acoustic transfer function of the vocal tract.

Before the subject makes an actual measurement, a continuous noise sequence is played out to the shaker. The subject can listen to the response through the headphones, and thus find the point on the neck that gives the loudest response. The duration of the actual measurement is only 150ms, allowing the subject to maintain a constant vocal tract shape without much difficulty. Previous methods of measuring the VT transfer function had the disadvantage that a large measurement time was required. The sweeptone system of [Fujimura and Lindqvist 1971] required the subject to maintain a constant vocal tract shape for 8.5 secs, and training was necessary.

The transfer function is calculated using correlation techniques. The particular properties of the pseudo-random excitation sequence, and the method of analysis shown in section 4.2, mean that external signals (signals not caused by the excitation sequence) picked up by the microphone are largely *correlated-out* of the final result.

A novel method of using the system to measure the closed tract transfer function of the vocal tract is also presented. The closed tract resonance is a useful parameter in the quantification of energy loss in the vocal tract. The new method can be used as an alternative, or a supplement to the impedance measurements presented in Chapter 6.

4.2 Summary of the theoretical background

The vocal tract is considered as a linear filter, with an impulse response h(n), and a transfer function H(f). The vocal tract is artificially excited with a known noise sequence x(n). Natural excitation, such as speech sound production, is represented by the term b(n). The response of the vocal tract to the combined natural and artificial excitation can be represented as:

$$y(n) = h(n) * [b(n) + x(n)] \tag{4.1}$$

a convolution sum given by:

$$y(n) = \sum_{k=-\infty}^{\infty} h(k)[x(n-k) + b(n-k)] \tag{4.2}$$

The cross correlation $R_{xy}(n)$ between the sequences x(n) and y(n) is defined as:

$$R_{xy}(n) = \sum_{m=-\infty}^{\infty} x(m)y(m+n) \tag{4.3}$$

Substituting the equation 4.2 into equation 4.3 we can see that:

$$R_{xy}(n) = R_1(n) + R_2(n) \tag{4.4}$$

$$R_1(n) = \sum_{k=-\infty}^{\infty} h(k)\left[\sum_{m=-\infty}^{\infty} x(m).x(m+n-k)\right] \tag{4.5}$$

$$R_2(n) = \sum_{k=-\infty}^{\infty} h(k)\left[\sum_{m=-\infty}^{\infty} x(m).b(m+n-k)\right] \tag{4.6}$$

Let $\phi_{xx}(n)$ be the autocorrelation of x(n), and $\phi_{xb}(n)$ the correlation between x(n) and b(n). Then:

$$R_{xy}(n) = [h * \phi_{xx}](n) + [h * \phi_{xb}](n) \tag{4.7}$$

Assuming that x(n) and b(n) are uncorrelated then $\phi_{xb}(n) = 0$. This implies:

$$R_{xy}(n) = [h * \phi_{xx}](n) \tag{4.8}$$

The closer $\phi_{xx}(n)$ is to a Dirac impulse then the nearer $R_{xy}(n)$ will be to h(n). It is clear that for accurate estimation of h(n) we need an input seqeunce x(n) that exhibits a sharp peak as its autocorrelation. The pseudo-random or Galois sequences satisfy this autocorrelation requirement. The circular autocorrelation of a pseudo-random sequence of length $N = 2^m - 1$ has the following characteristic [Schroeder 1984].

$$c_n = \begin{cases} 2^m - 1 & \text{for } n \equiv 0 \mod(2^m - 1) \\ -1 & \text{otherwise} \end{cases} \tag{4.9}$$

The pseudo-random sequences are generated using maximal length feedback shift registers as shown in Chapter 3. We notice here that this condition applies only to a *circular* autocorrelation of the pseudo-random sequence as opposed to a *linear* auto-correlation. The above analysis is now re-worked for the case of circular correlations, and takes account of the fact that finite length sequences are used. The circular autocorrelation of a finite length sequence of length N is defined as:

$$\tilde{\phi}_{xx}(n) = \sum_{m=0}^{N-1} x(m)x(m+n, \mod(N)) \qquad n = 0,1,\cdots,N-1 \tag{4.10}$$

Similarly the circular crosscorrelation (as opposed to linear crosscorrelation) of two finite length sequences of length N, is defined as:

$$\tilde{r}_{xy}(n) = \sum_{m=0}^{N-1} x(m)y(m+n, \text{mod}(N)) \qquad n = 0,1,\cdots,N-1 \qquad (4.11)$$

It can be shown that for finite length sequences and when circular correlation is used, equation 4.8 above becomes:

$$\tilde{r}_{xy}(n) = \left[h * \tilde{\phi}_{xx}\right](n) \qquad (4.12)$$

If the length of the impulse response h(n) is smaller than N, and $\tilde{\phi}_{xx}(n)$ is a Dirac impulse the sequence $\tilde{r}_{xy}(n)$ corresponds exactly to h(n) for n varying from 0 to N-1. In this case if $FT\left[\tilde{r}_{xy}\right](k)$ is the Fourier transform of length N of $\tilde{r}_{xy}(n)$ we get:

$$FT\left[\tilde{r}_{xy}\right](k) = H(k) \qquad (4.13)$$

Let $\tilde{\Phi}_{xx}(k)$ be the discrete Fourier transform of $\tilde{\phi}_{xx}(n)$:

$$FT\left[\tilde{r}_{xy}\right](k) = H(k).\tilde{\Phi}_{xx}(k) \qquad (4.14)$$

The Fourier transform of a peak is flat, or a constant so $\tilde{\Phi}_{xx}(k)$ is a constant $\tilde{\Phi}_0$

Thus: $$FT\left[\tilde{r}_{xy}\right](k) = H(k).\tilde{\Phi}_0 \qquad (4.15)$$

Equation 4.15 shows how the transfer function H(k) can be determined from the circular cross-correlation of the measured sequences. The length of the pseudo-random sequence was chosen as 2047 samples. The sampling rate used was 16.348kHz. The sequence was zero-padded to length 2400 samples to take account of the delay between the input and output, and to allow some settling time. The actual measurement time is 147ms. The block denoted 'Vocal Tract' in the figure 4.1 overleaf also includes the transfer function of everything in the signal path from the D/A to the A/D converters. This includes the transfer function of the shaker, and the transfer function of the neck wall. The transfer function of the neck wall is not known, but in general is assumed to have a low-pass filtering effect with a fairly high cutoff frequency [Djeradi et al, 1991]. The shaker (Ling Dynamic Systems Model V201) has a reasonably flat response up to 13kHz. In the future it may be tested under skin load conditions using an accelerometer.

4.3 Experimental approach and measurement procedure

The software and hardware used for the experiment were as follows:

Software: An I/O system that simultaneously records and plays back. A control program that performs housekeeping and mathematical manipulation of the results.

Hardware: An Electromagnetic Shaker, an LSI C30 board, a Bruel&Kjaer condenser microphone, two amplifiers and two analog lowpass filters.

The I/O system (see chapter 3) was implemented on a LSI TMS320C30 based development platform. A serial port analog interface to the DSP card was designed and built to increase the A/D and D/A conversion accuracy of the system, and also to allow more flexible two channel sampling. An improvement in accuracy from 10 bits to 13 bits was obtained (see Chapter 3). Before the measurement system was applied to the human vocal tract some test measurements were carried out.

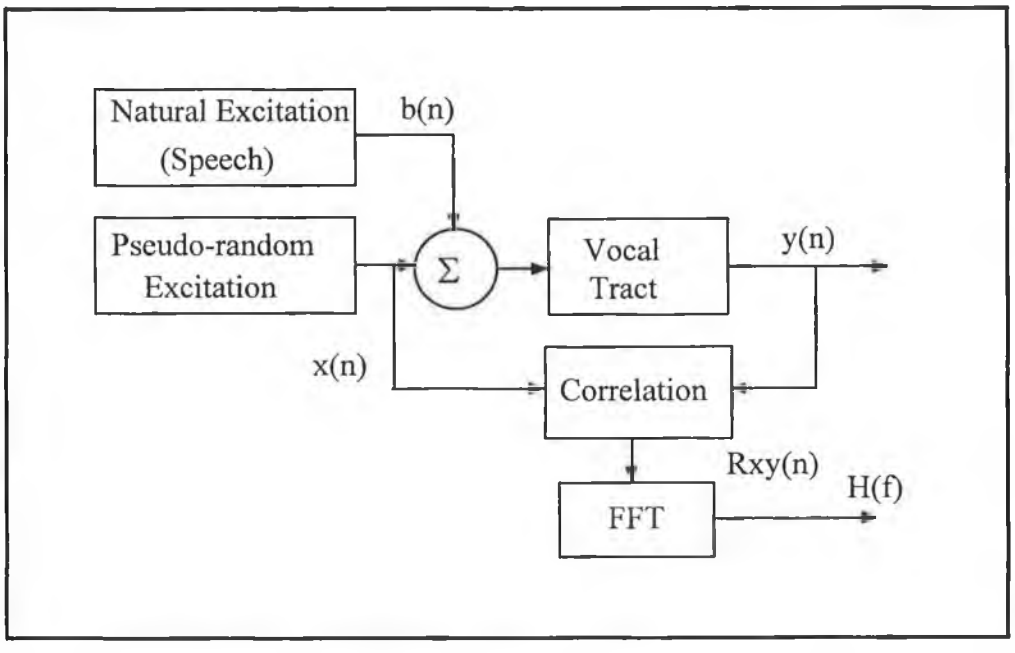


Figure 4.1: Principle of a vocal tract transfer function measurement.

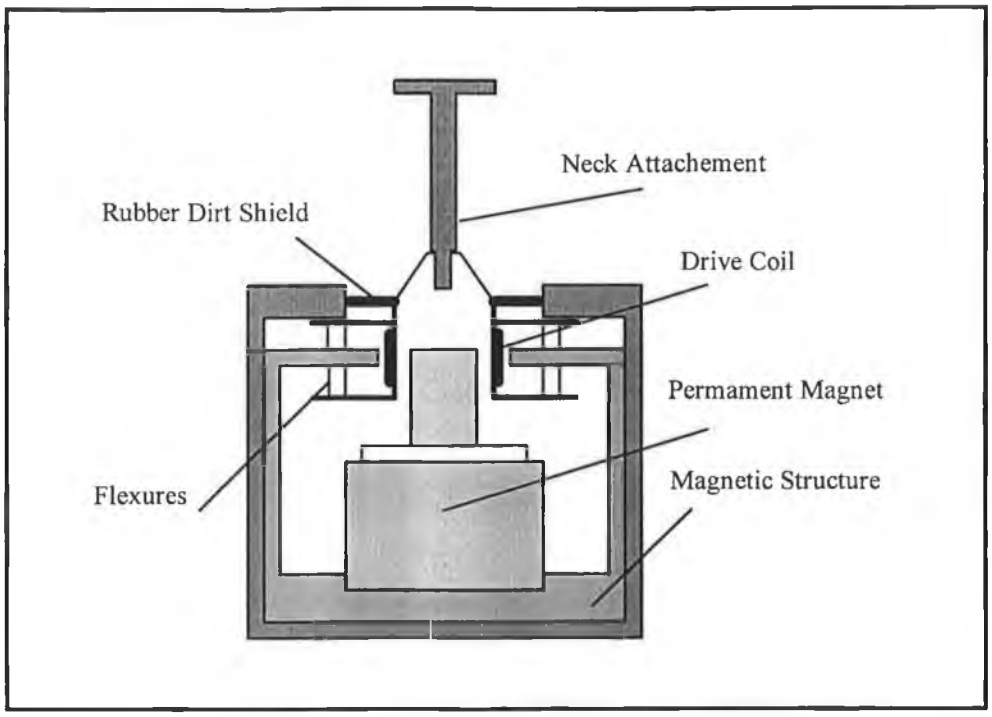
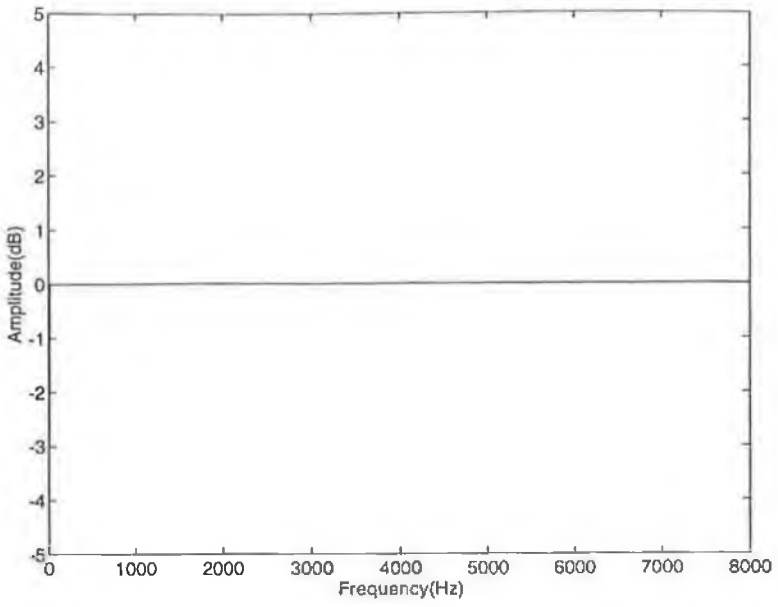


Figure 4.2: Sectional drawing of the shaker with neck attachement

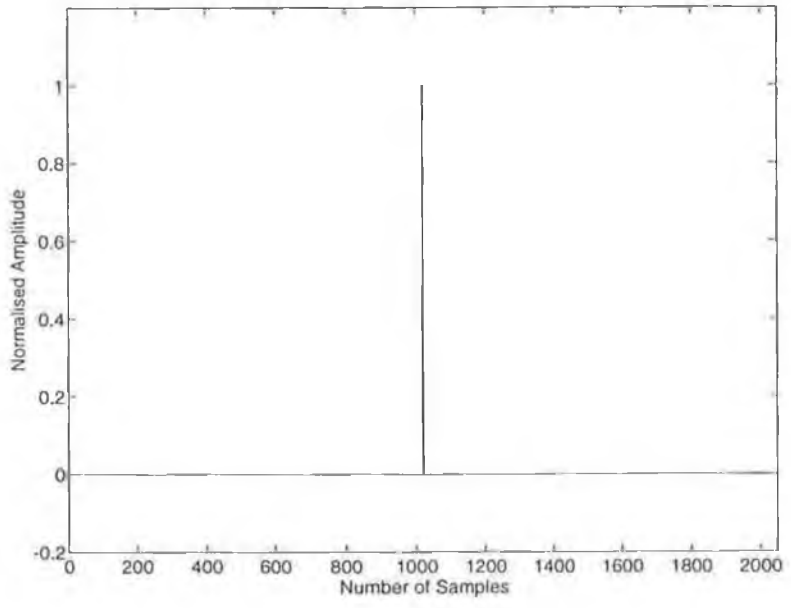
The first measurement was to feed the output of the system directly to the input and perform a cross-correlation. Essentially this means removing the block labelled 'vocal tract' from the block diagram shown in figure 4.1. In this case the recorded input sequence should be exactly the same as the output sequence except for additive external noise effects. The cross-correlation of the input and output, then becomes an autocorrelation of the input sequence. This, as shown in the section 4.1, should be a dirac-delta function. The FFT of this function should then be flat, with any deviation being due to system noise.

Figure 4.3 overleaf shows the results of the direct feedthrough measurement in both the time and the frequency domain. The results are as expected with very little effects from system noise.

Next, the system was used to measure the electrical transfer function of the analog 5kHz lowpass filter used in the measurement. This was to ensure that the filter had a reasonably flat response in the range 0-5kHz, and also to show that the system worked (see figure 4.4 overleaf)



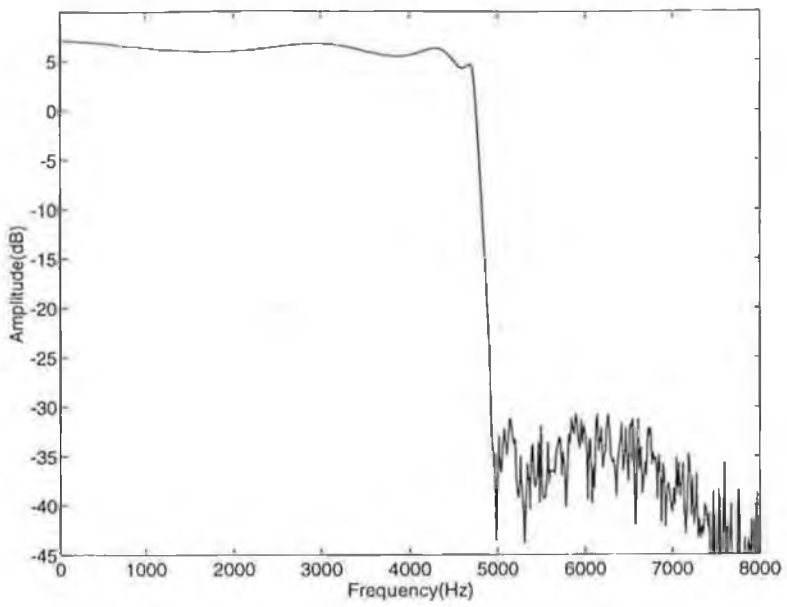
(a)



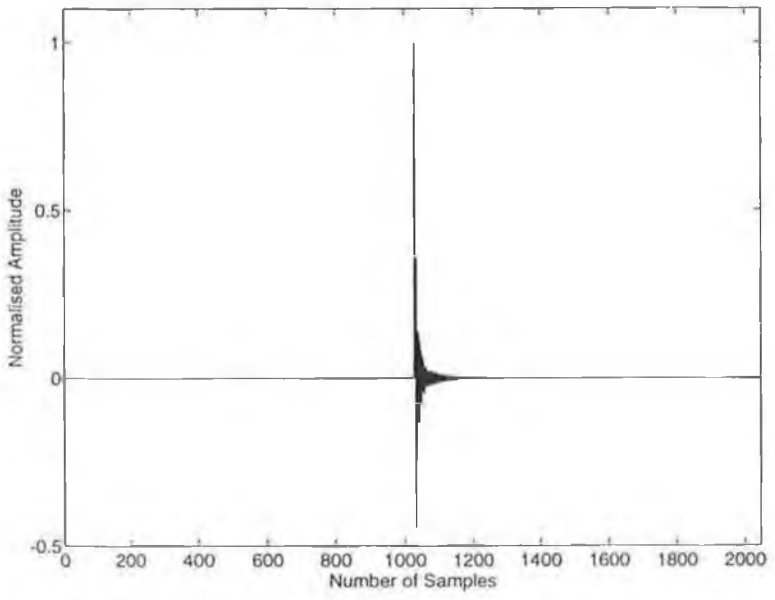
(b)

Figure 4.3: (a) Measured transfer function of the system for direct feedthrough from output to input

(b) Measured cross-correlation peak for direct feedthrough from output to input



(a)



(b)

Figure 4.4: (a) Measured transfer function of a 5kHz analog filter.

(b) Measured impulse response of a 5kHz analog filter.

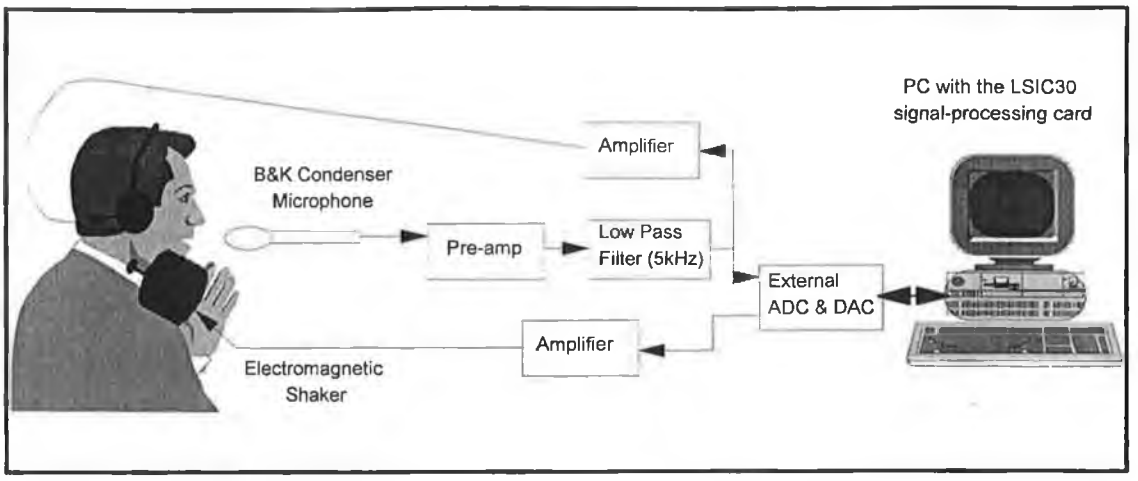


Figure 4.5: Experimental Setup for Speech Measurements

4.4 Speech measurements

Closed glottis vowel measurements were performed in DCU's semi-anechoic studio. An attachment of a circular disk on a thin metal shaft was made for the shaker so that measurements could be performed on the human vocal tract (see figure 4.2). The subject presses the shaker to the throat at the level of the thyroid cartilage. Simultaneously the response is measured at about 1cm from the lips using a condenser microphone (Bruel & Kjaer type 4006). A wooden baffle with a cutout for the subjects neck is placed between the shaker and the microphone to minimise direct radiation. It is very important to keep this direct radiation as low as possible, otherwise interference will occur with visible effects. Tests were performed on the baffle with the shaker and the microphone in measurement position, but without the shaker pressed against the subject's neck. The measurements showed that best results are obtained if the excitation level is kept low, and the microphone amplification is high. In this case the level of direct radiation from the shaker recorded by the microphone is almost as low as background noise in the room.

Initially, a continuous noise sequence is outputted to the shaker allowing the subject to listen to the amplified microphone response through headphones, and find the point on the neck that gives the loudest response. The recorded signal is filtered using a 5kHz analog lowpass filter. While a measurement is being taken the subject articulates the vowel silently while attempting to keep the glottis closed and the velum raised. The computed transfer function is immediately displayed on a PC screen

placed outside the studio window, allowing the subject to check the results. For a batch of measurements the program may be run without viewing the results.

The measurements discussed below are of non-nasalised vowels in closed glottis conditions. For these measurements the glottis was kept closed to avoid coupling with the subglottal cavities, and the velum was raised to avoid shunting in the nasal tract as a side branch.

4.5 Results

The results below (figures 4.6 to 4.10) show the measured transfer functions (in closed glottis conditions) for the vowels [a], [ɛ], [i], [o], and [u]. Very different articulator positions are represented here, and this is evident from the differences in transfer function for the vowels. The formant peaks are clearly visible for frequencies less than 4kHz. The sharp zero around 4.5 kHz is typical of this type of measurement, and is due to the fact that the point of excitation is not exactly at the closed end of the VT (the glottis) but is slightly above it. For instance, if the shaker is placed at distance x above the glottis, we would expect zeros at frequencies F_n= n.c/2.x where c is the speed of sound in air and n is an integer. For a shaker height of 4cm, the first zero should occur F_1=4.2kHz. If the shaker is placed too high on the neck, this zero decreases in frequency and may cancel one of the VT resonances.

The results shown are for one subject [BH], and it can be seen that a high degree of repeatability is obtained. The subject was a male adult, 178cm in height and weighing approximately 71 Kg. The subjects background is middle class Irish, with a relatively neutral accent. Variation in formant frequencies for the same vowel are mainly due to small differences in the position of the articulators between measurements. The formant frequencies and bandwidths are extracted automatically from the data. The frequencies are accurate to within 2Hz and the bandwidths to within 4Hz. A typical set of results for each of the measured vowels is shown in table 4.1.

	[a]	[ɛ]	[i]	[o]	[u]
F1	574	531	271	415	261
F2	1173	1565	2183	976	906
F3	2608	2444	2748	2500	2177
F4	3296	3436	3520	3125	3292
B1	52	44	66	48	54
B2	58	84	48	26	36
B3	62	94	184	38	38
B4	78	122	166	62	62

Table 4.1: The first four Formants and Bandwidths (in Hz) measured for five vowels.

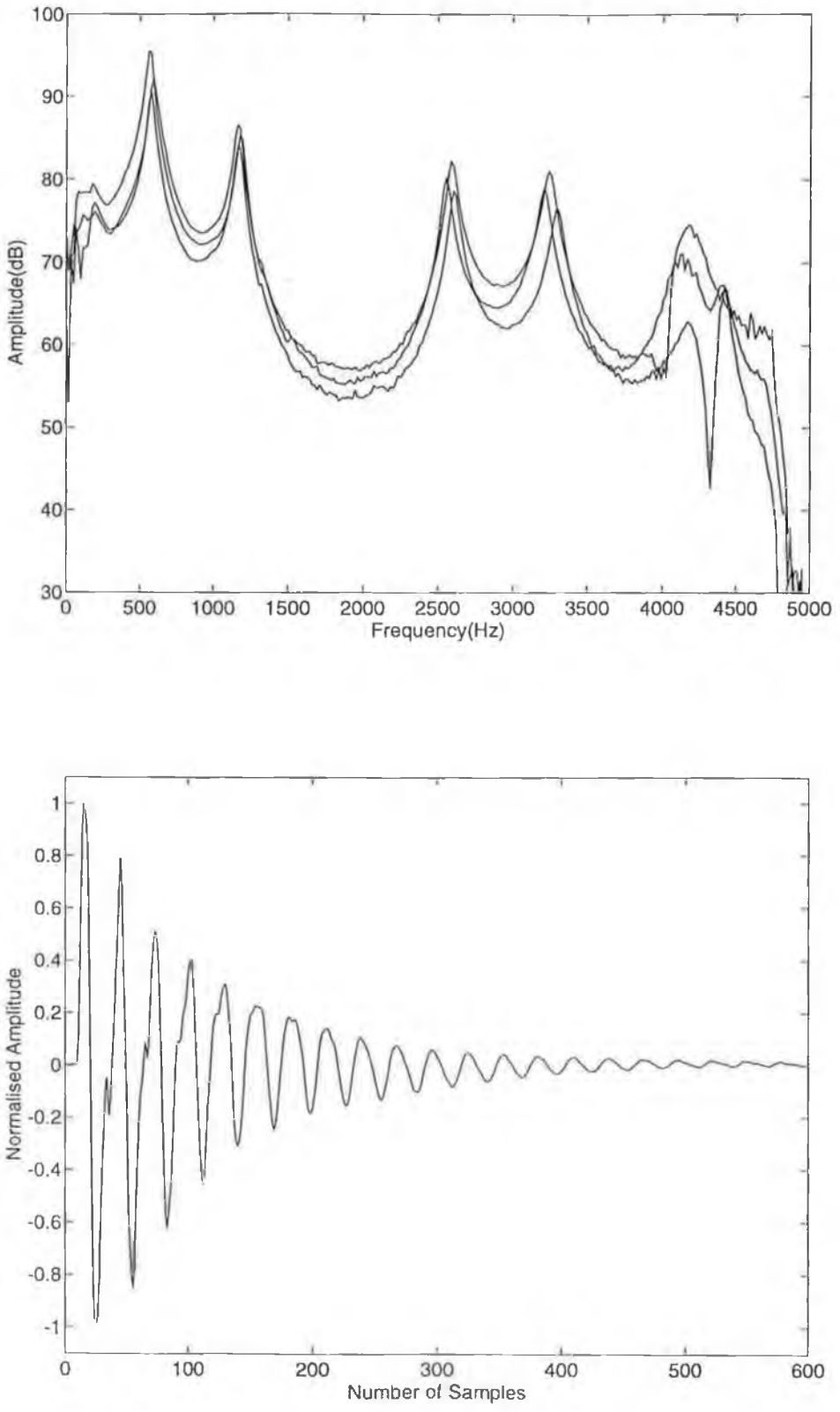


Fig. 4.6 : Measured Acoustic Transfer Function (top) and Impulse Response of the Vocal Tract for the vowel [a]

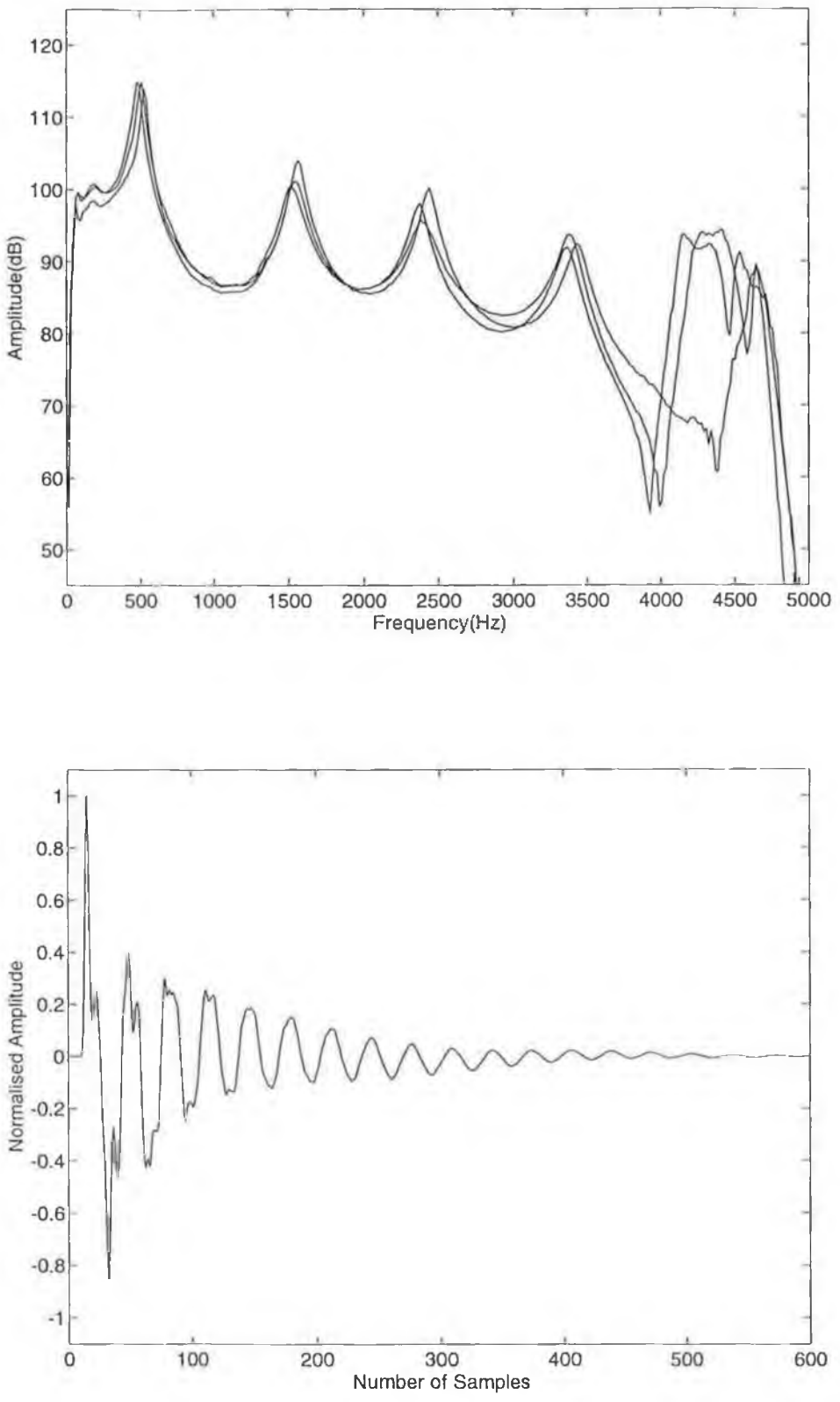


Fig. 4.7 : Measured Acoustic Transfer Function (top) and Impulse Response of the Vocal Tract for the vowel [ε]

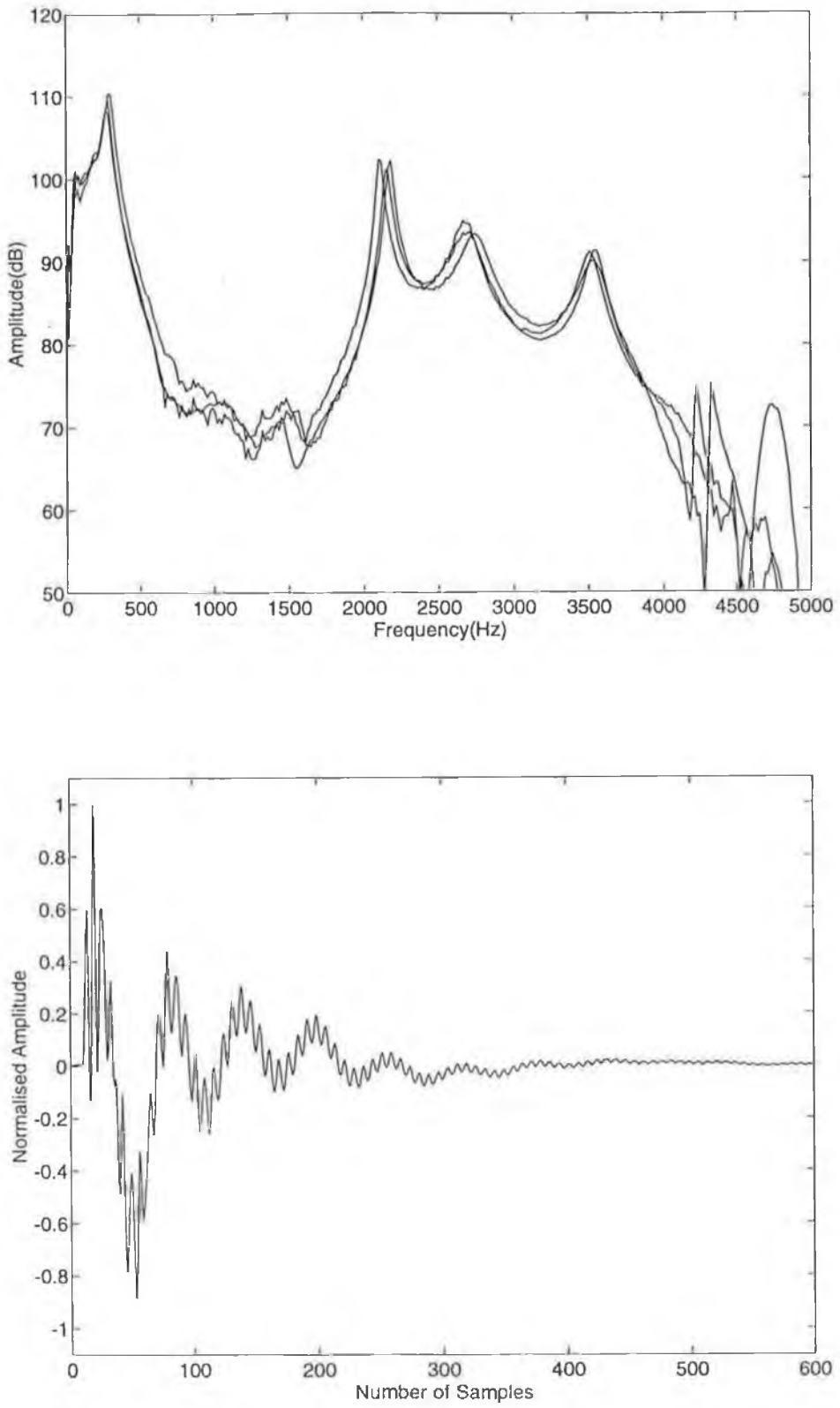


Fig. 4.8 : Measured Acoustic Transfer Function (top) and Impulse Response of the Vocal Tract for the vowel [i]

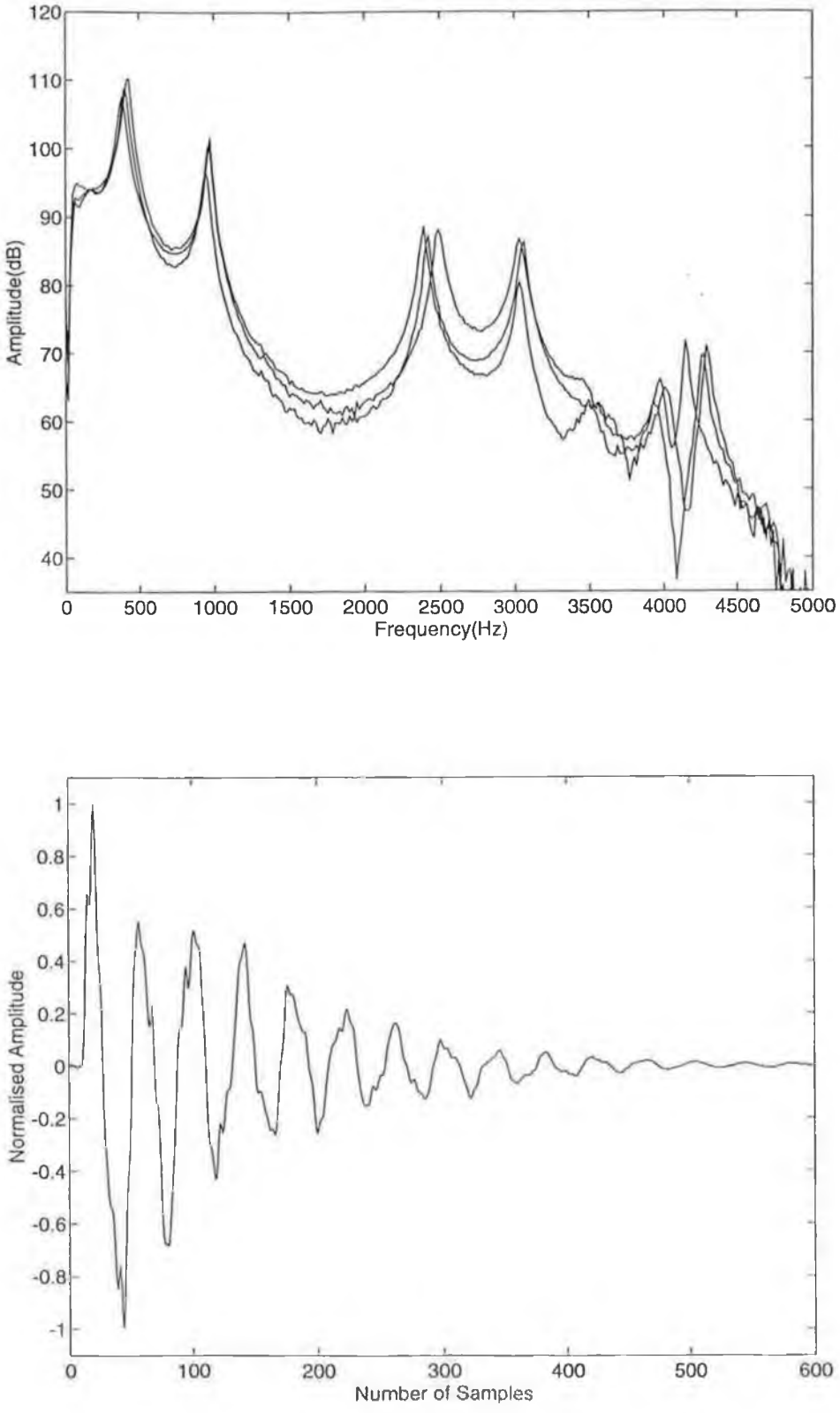


Fig. 4.9 : Measured Acoustic Transfer Function (top) and Impulse Response of the Vocal Tract for the vowel [o]

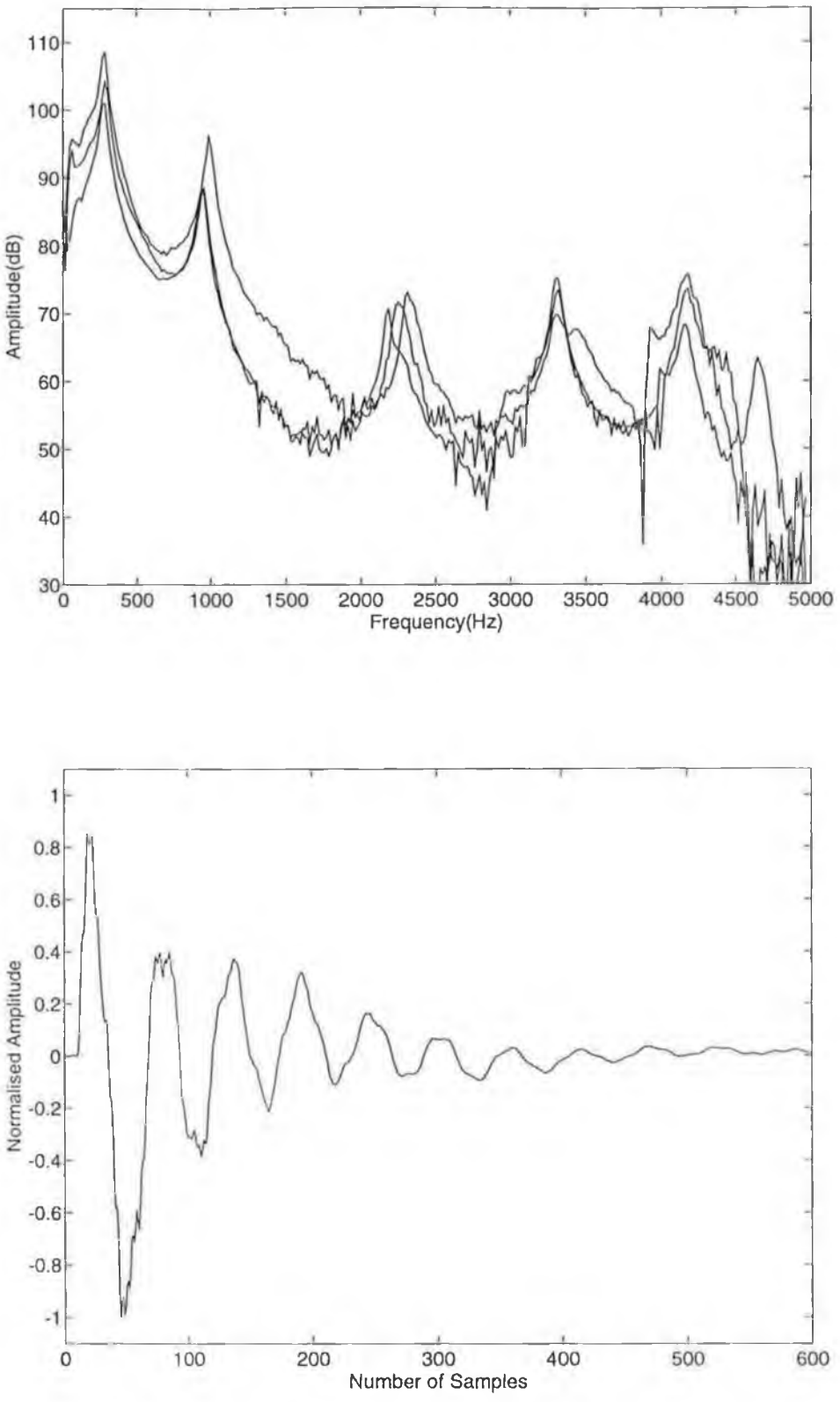


Fig. 4.10 : Measured Acoustic Transfer Function (top) and Impulse Response of the Vocal Tract for the vowel [u]

4.6 The Maeda model of the vocal tract system

Maeda [Maeda 1982] described software for a digital simulation of the vocal tract. This is based on a numerical solution of the acoustic transmission line equations relating the pressure p(x,t) and the volume velocity u(x,t) inside an acoustic tube. Equations are included for an estimation of the movement of the vocal tract walls based on a specification of mass per unit length along the vocal tract. The vocal tract is generally specified by its area function, that is the cross-sectional area of the tract as a function of distance along the tract from the glottis to the lips.

Maeda's model allows simulation of a vocal tract acoustic transfer function from a given area function. Fant [Fant 1960] and others have derived area functions for various utterances from the analysis of x-ray measurements, and their data may be used as input to Maeda's model. Maeda has also developed a 7-parameter articulatory model of the vocal tract based on a statistical analysis of area function data [Maeda 1979]. The parameters of this model include tongue shape, jaw position, lip height etc. Individual shapes may be stored using only 7 parameters. The model will calculate an area-function from these parameters, and again the transfer function may be obtained. The area-function calculated from the articulatory parameters may be slightly different from an actual measured area function, and thus transfer functions calculated from the two will also be different.

The transfer functions generated by Maeda's model for utterances similar (but not exactly the same) to those measured are shown overleaf (see figures 4.12). A combination of the above methods is used, as Fant's data is based on a Russian speaker [Fant 1972], and the data used by Maeda is based on French speakers, and so only a qualitative assessment is possible. For [a] the area-function used was from Fant's measurements, for [i] and [o] the area function data was from the French speakers, and for [ɛ] and [u] the articulatory model was used.

The transfer functions produced by Maeda's model are the ratio of volume velocity at the lips and nostrils, to the volume velocity at the glottis, wheras the measured transfer functions are the ratio of pressure at the lips and nostrils to the volume velocity at the glottis. The pressure at the lips and the volume velocity at the lips are related by the acoustic radiation impedance at the lips $Z_L(f)$.

$$P(l, f) = Z_L(f).U(l, f) \quad (4.16)$$

Where l is the distance from the glottis to the lips. Maeda takes account of $Z_L(f)$ as a load on the output of his model. The load he assumes is based on Flanagan's [Flanagan 1972] approximation of the radiation impedance as a resistor and an inductor in parallel. For the above utterances the nasal tract is bypassed so the pressure and volume velocity at the nostrils are zero. It follows that the two transfer functions are also related by $Z_L(f)$ ie.

$$H_{measured}(f) = Z_L(f).H_{modelled}(f) \quad (4.17)$$

To a first approximation the overall effect of this impedance function may be assumed to be a boost in the measured transfer function of +6dB per octave.

The purpose of this qualitative comparison of the measured results with the modelled results for the same vowel (although from different speakers) is mainly to show the similarity in the overall spectral shapes. This at least shows that the measurements are on course. The modelled transfer functions can be used to generate the vowel sounds on a PC to subjectively test their validity.

A more quantitative comparison of the matching between the measured results and the model can be attempted if the area-function data used as input for the model is also known for the subject. The spectral roll-off of the measured transfer functions is greater than the modelled transfer function despite the radiation boost. This is due to the undetermined transfer function between the electrical signal exciting the shaker and the volume velocity of air generated in the laryngeal area. It is expected that this would be a second order lowpass filter [Badin et al 1993], with a roll-off of approximately -12dB/octave. This is supported by the fact that the overall slope of the measured results is -6dB/octave.

4.7 The closed tract resonance.

It is generally agreed that existing articulatory models could be improved by more accurate modelling of energy loss mechanisms in the vocal tract. In real speech, formant frequencies and bandwidths are directly affected by these energy loss mechanisms. It is well known that the bandwidths of the lowest two formants depend primarily on the vocal tract wall losses, whereas the bandwidths of the higher formants

depend primarily on lip radiation loss, and on viscous and thermal losses in the VT. Most energy loss occurs through vibration of the non-rigid vocal tract walls at low frequencies. The first formant decreases in frequency as the lip opening becomes smaller. The extreme lower limit of F_1 is the closed tract resonance frequency denoted by F_W (with bandwidth B_W). Fant [Fant 1976] has modelled this resonance using an equivalent circuit of a resistor R_w and inductor L_W connected in parallel with a capacitor C_W (see section 5.6 and figure 5.6, where this circuit is discussed in detail). C_W is related to the volume of air in the tract, and R_W and L_W are related to the impedance of the VT walls. He has shown that:

$$F_W = \frac{1}{2\pi}(L_W C_W)^{-\frac{1}{2}} \tag{4.18}$$

$$B_W = \frac{1}{2\pi} \cdot \frac{R_W}{L_W} \tag{4.19}$$

Measurements of this resonance for different speakers can be useful in helping to quantify wall losses. The closed tract response can be measured using the shaker apparatus. The lips are sealed tightly around a microphone and the VT is held in a neutral position. As before, the glottis is closed and the velum is raised. The transfer function generated is over the same frequency range as before (0-5kHz), and from this the first resonance may be extracted. Initial measurements have shown values of $F_w = 180\text{Hz}$ and $B_w = 75\text{Hz}$, in agreement with Fant's measurements. A plot of a typical measurement is shown below.

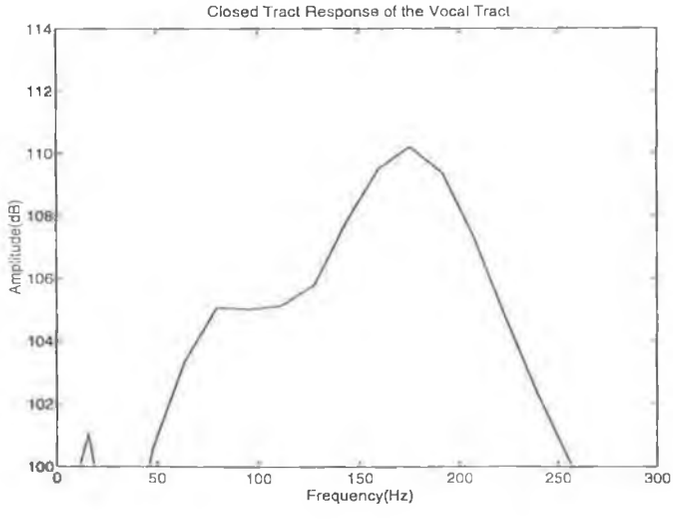


Figure 4.11: The first closed tract resonance

4.8 Summary

This chapter describes an implementation of the Electromagnetic Shaker experiment of Djeradi et al, for measurement of the acoustic transfer function of the vocal tract. Speech measurements have yielded results similar to those presented in the Djeradi paper. Area-function data, and articulatory parameters have been applied to a digital model of the vocal tract and the generated transfer functions have been compared to the measured data. A good qualitative match of formant position is observed for the first four formants, despite the fact that the data is for different speakers, which at least indicates that the measurements are in the correct region. A novel method of applying the method to the study of vocal tract energy loss mechanisms has been discussed. A system for dual transfer-function/impedance measurements was also implemented and this system is discussed in the next chapter.

[a]

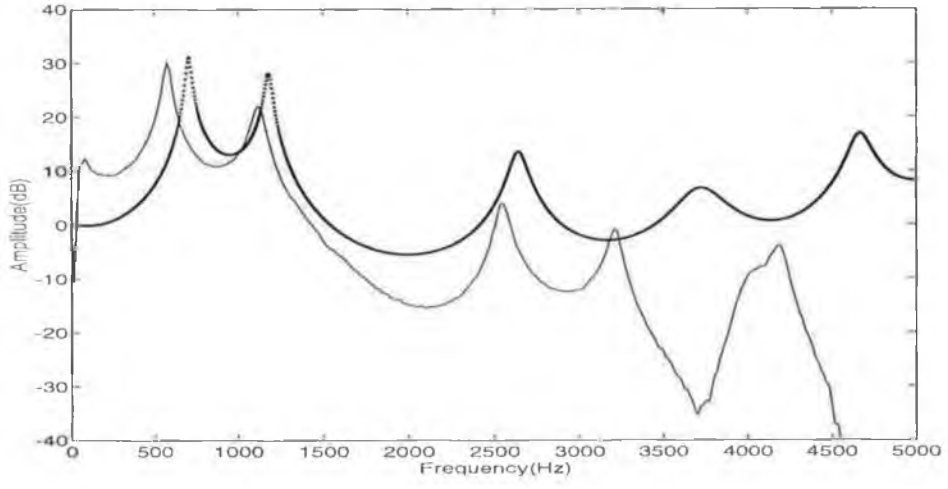
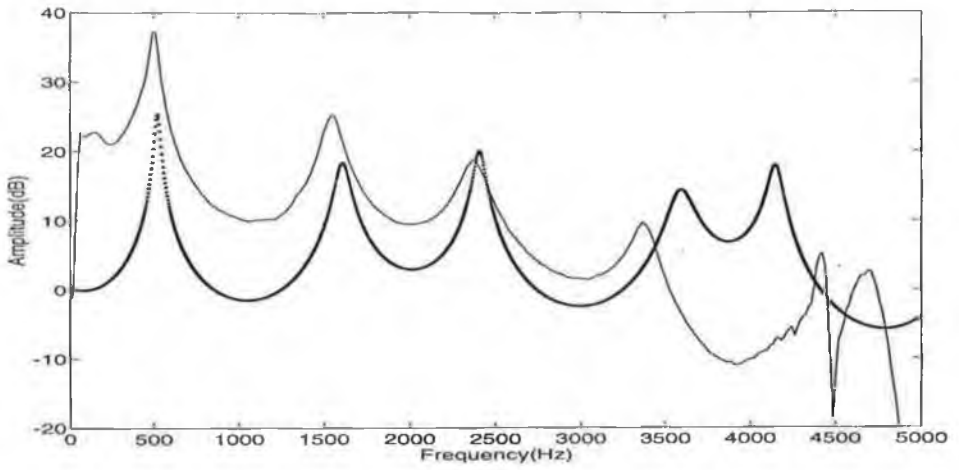


Fig. 4.12a: Qualitative comparison of the measured (solid line) and modelled transfer functions

[ε]



[i]

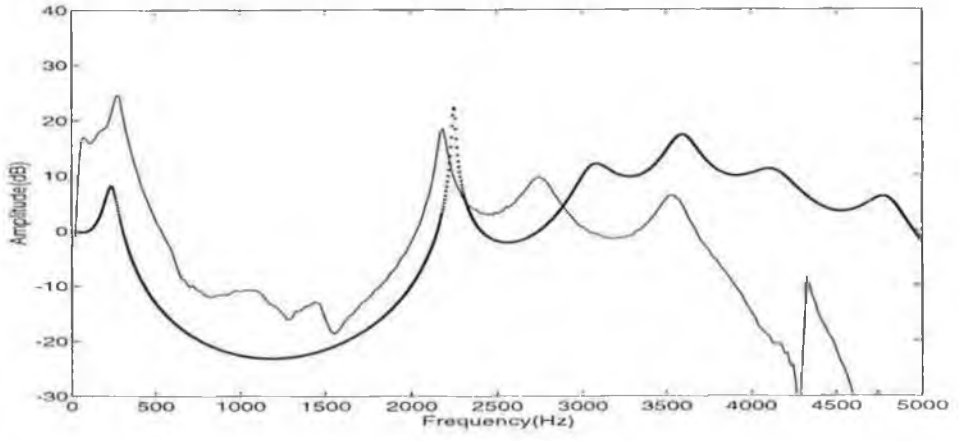
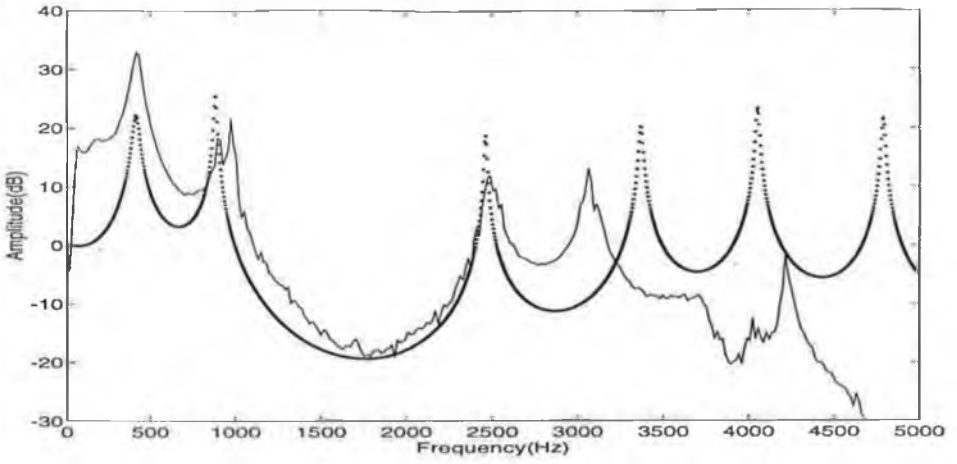


Fig. 4.12b: Qualitative comparison of the measured (solid line) and modelled transfer functions

[o]



[u]

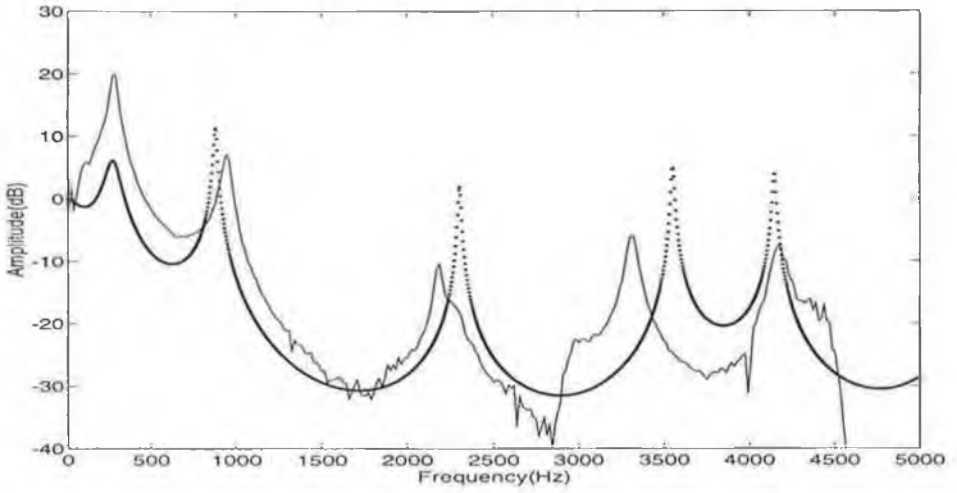


Fig. 4.12c: Qualitative comparison of the measured (solid line) and modelled transfer functions

Chapter 5

Quasi-Simultaneous Transfer Function & Acoustic Impedance Measurements

5.1 Introduction

The main aim of this thesis is to investigate the variation of the wall vibration load in a wide range of human subjects. This aim is prompted by the answer to the larger question: Is the wall vibration load important in the study and implementation of vocal tract and articulatory models? We show here that it is agreed that the answer to this question is unequivocally yes. The wall vibration load has a fundamental effect on the quality of our speech. This then leads to a number of following questions, including: How can we add the effect of the wall vibration load to our acoustic and articulatory models? Can the wall vibration load or an estimate to the wall vibration load be measured in real human subjects, and if so, how does the wall vibration load vary between males and females, between adults and children?

The difficulty of performing acoustic measurements in the past meant that some of these questions have been left unanswered, and even uninvestigated. This leaves those acoustic and articulatory modellers in the position of trying to improve their models without any knowledge of how these effects work in real life. As the whole idea of articulatory models is that they attempt to represent what is actually occurring in real life, it is clear that gaps in our knowledge need to be filled.

In this chapter some of the above questions are tackled. First of all the importance of the wall vibration load and methods for its measurement are discussed. Next the methods used in DCU to build and calibrate a wall vibration load measurement system are summarised, with detailed descriptions given in appendix B.

In order to combine the knowledge gained from transfer function measurements and measurements of the wall vibration load, a measurement system to perform quasi-simultaneous transfer function/acoustic input impedance measurements was

developed. This system allows a transfer function measurement to be made, closely followed by an acoustic input impedance measurement. While a full scale survey was not made using the system, the system was designed, developed and measurements were performed on trained subjects. The operation of the rig to perform dual measurements is described.

5.2 The importance of the wall vibration load

The wall vibration load is known to affect the tuning of low frequency F1, and also pharyngeally constricted sounds (such as [a]) with high F1. It has a fundamental effect on the quality of our speech [Fant, 1970]. The wall vibration load has primarily a low frequency effect. The closed tract resonance frequency is a natural low-frequency resonance of the vocal tract and is due to the synergy between the vocal tract walls and the volume of the air contained in the vocal tract. Measurements of the closed tract resonance can be used to obtain information about the wall vibration load.

Vocal tract lip impedance measurements were originally performed by Fant, Branderud and Nord (1976) to obtain information about the wall vibration load of the vocal tract. The method used by Fant et al to measure parameters of the wall vibration load was based on injection of a sweep-tone into the subject's vocal tract using a narrow plastic probe tube inserted in the lips. The response was measured via a pick-up probe tube also inserted in the lips and connected to a microphone. The frequency range of the measurements was 65Hz-650Hz and the duration of the measurement was 1s. Their experiments showed a first resonance peak at around 190Hz for the average male, and at around 220Hz for the average female. Using an equivalent circuit model of the wall load, they were able to extract equivalent circuit component values from measurements of this resonance. The circuit component values can be used directly as parameters in a lumped wall load section of a vocal tract model. The real mass load in the vocal tract is of course non-uniformly distributed over the entire length of the vocal tract. However, other measurements by Fant et al (1976) have shown that the major losses occur at the glottal and lip ends making the concept of lumping the load at these points realistic.

Can we show how these sort of lumped wall vibration load parameters will affect the output of a vocal tract simulator? Two of the wall vibration load equivalent circuit

components are L_w and R_w, the wall vibration load inductance and the wall vibration load resistance. These parameters are explained in section 5.6.

By applying measured values of these parameters to a vocal tract simulator, we can observe how the output of the simulator varies over the range of parameters.

The Maeda vocal tract simulator (see section 4.6) was used for an estimation of the variability of L_w and R_w on observable features in the vocal tract transfer function. The simulator was modified at DCU to allow for a single lumped wall load element 4cm above the glottis. Using values of L_w and R_w for subjects SC1 and DC2, at opposite extremes of the adult male group in the DCU wall vibration load survey (see chapter 6) and applying them to the area function corresponding to an [a] articulation, the following results were obtained.

Subject:	F1 (Hz)	B1 (Hz)	F2 (Hz)	B2 (Hz)	FZ1 (Hz)	BZ1 (Hz)
SC1	658.5	33.4	1536.5	45.9	264.0	59.8
DC2	611.0	27.5	1528.5	47.3	192.0	92.1
difference	-47.5	-5.9	-8.0	+1.4	-72.0	+32.3

Table 5.1: Effect of measured lumped loss components on the formants of a vocal tract simulator. FZ1 and BZ1 are the frequency and bandwidths of the first lip impedance peak.

As can be seen the lumped wall components have the greatest effect on F1. The difference of about 50Hz in F1 between the two cases is an example that the variation of these model parameters can have a significant effect upon the speech output, and as a result of the perceived speech quality of our models. Further work along this line in DCU will include computation of the sensitivity of the Maeda articulatory parameters to changes in the assumed values of L_w and R_w. This will be based on a multi-parameter optimisation procedure, using a linked formant-articulatory codebook to obtain suitable starting points.

5.3 Acoustic impedance measurements using the Salava device

In this chapter acoustic impedance measurements performed at DCU are discussed. The impedance measurement device was specially developed at DCU and its operation and calibration are described. With this device the frequency range of the measurements is from 100Hz to 5kHz and the duration of a measurement is 250ms. The extraction of wall vibration load information from impedance measurements on human subjects is explained.

5.3.1 The impedance measurement apparatus

The acoustic impedance device used in these measurements was developed by Scaife [1994] and was based on the work of Merhaut [1968] and Salava [1988]. Some measurements were also made with another device developed earlier by Burke and Scaife [1989]. The earlier device measured acoustic impedance by measuring the terminal electrical impedance of a loudspeaker acoustically coupled to the acoustic load to be measured. This device was successfully used to measure acoustic impedances of rigid tubes.

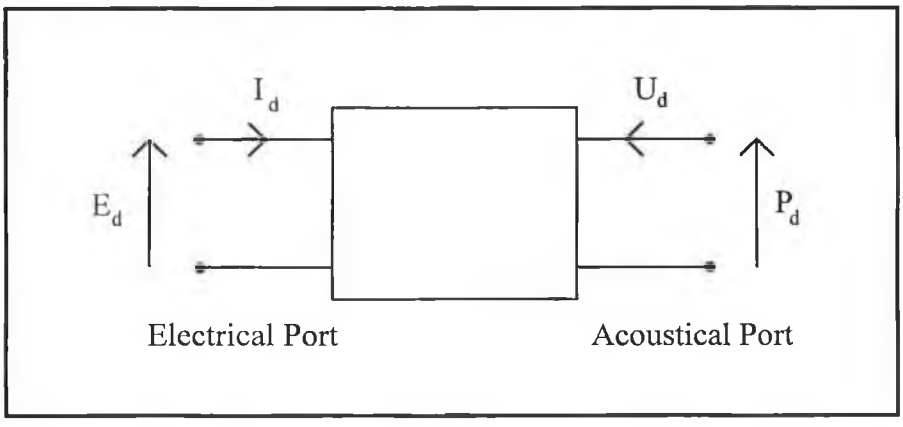


Figure 5.1: Electro-acoustical 2-port device

The earlier device was essentially a two-port system relating electrical impedance to acoustic impedance. The two ports of this system are related by the system transfer matrix. In this case, the relationship between the voltage E_d and current I_d into the electrical port, and the pressure P_d and volume velocity U_d into the acoustic port is given by:

$$\begin{pmatrix} E_d \\ I_d \end{pmatrix} = \begin{pmatrix} A_d & B_d \\ C_d & D_d \end{pmatrix} \cdot \begin{pmatrix} P_d \\ -U_d \end{pmatrix} \qquad (5.1)$$

Once the elements of the transfer matrix are known, knowledge of the signal at one of the ports allows determination of the signal at the other. For this system the acoustic impedance seen at the acoustic port could by determined from the electrical impedance measured at the electrical port by:

$$Z_a = \frac{B_d - D_d . Z_e}{C_d . Z_e - A_d} \qquad (5.2)$$

The four elements of the system transfer matrix (A_d, B_d, C_d, D_d) were determined by a calibration process whereby a known electrical signal was applied to various acoustic loads. Experiments with this device showed that while it successfully measured the input acoustic impedance of rigid tubes, it was unsuitable for measurements on human subjects due to distortion caused by temperature and pressure effects with real subjects.

5.3.2 The Salava device

The second device (hereafter referred to as the Salava device) proved much more successful in the measurement of acoustic impedance for both rigid tubes and human subjects. The measurement device is based on the fact that acoustic impedance at a point in air is defined as the frequency domain ratio of the acoustic pressure to the volume velocity of the air at that point.

$$Z(j\omega) = \frac{P(j\omega)}{U(j\omega)} \qquad (5.3)$$

Normally the pressure is easily measured using a microphone, but the volume velocity is quite difficult to measure (acoustic flow meters are bulky and expensive). With the Salava device (see fig 5.2) both of these quantities are directly measured near the end of a 25mm diameter copper tube which is coupled to the acoustic load to be measured.

A Sony MDR 'Walkman' headphone loudspeaker is attached at one end of the tube and is acoustically driven from behind by a 75mm mylar cone wide-range loudspeaker. The coupling between the two is improved by filling the space between them with modelling clay. The coupling drops rapidly above 2kHz which is the frequency of resonance between the masses of the loudspeakers and the compliance of the intervening cavity.

The headphone loudspeaker acts as a velocity microphone, Its open-circuit terminal voltage being proportional to the linear velocity of the diaphragm. The pressure is measured at a point 8mm in front of the headphone diaphragm using a 6mm electret microphone inserted into the copper tube. This spacing was recommended by Keefe and Benade [1981], to minimise the amplitude of evanescent or non-propagating modes from the diaphragm picked up by the microphone.

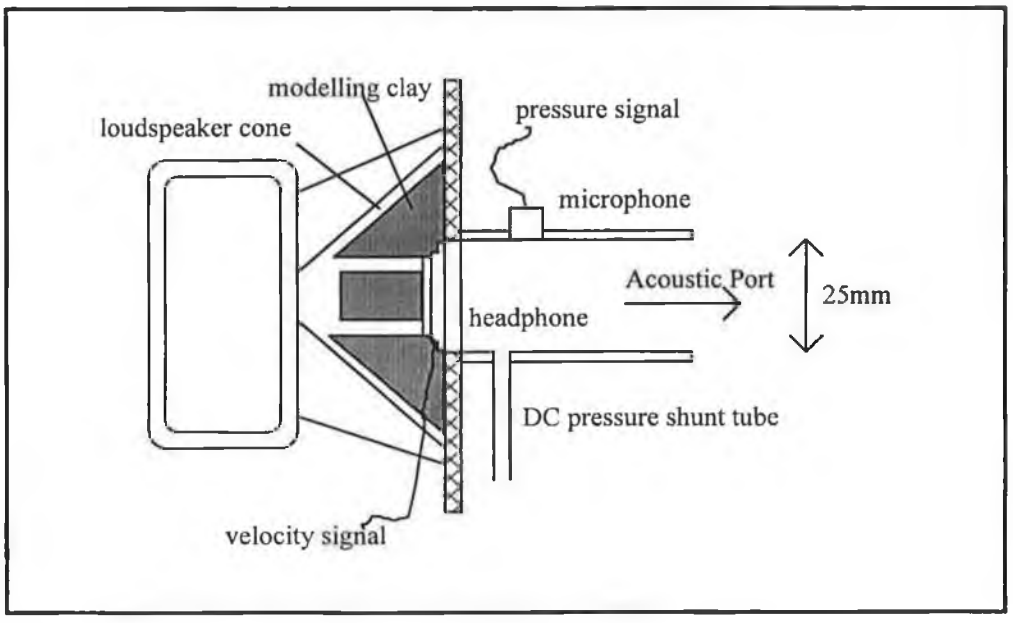


Figure 5.2: The impedance measurement device (The Salava device)

The microphone has a frequency response that is essentially flat from 50Hz up to 3kHz, with a small broad resonance peak of about 1dB at 3.5kHz. Evidence from the calibration process, described below, suggests that the e.m.f from the headphone is almost exactly proportional to the net flow in front of its diaphragm over the range of frequencies at which the system is used (50-4000Hz), in other words the diaphragm moves essentially as a rigid piston over this frequency range. This is reasonable as the headphone diaphragm consists of a rigid mylar dome inside the voice coil, with a narrow flexible outer suspension.

A pressure equalisation shunt tube was added to avoid displacement of the diaphragm by breathing. The subjects are not supposed to breath during the measurements, but it is inevitable that there will be some movement of air between the subject and the coupling tube. The dimensions of the shunt were carefully chosen based on two opposing requirements. The first is that the DC flow resistance of the tube is not too

high, otherwise it will not shunt DC pressure from the diaphragm adequately. The second that the AC impedance of the shunt tube is not too low, otherwise the measured impedance results would be inaccurate. The dimensions chosen of length 60mm and diameter 2mm give a DC flow resistance of 28cgs ohms and an AC impedance of 150 cgs ohms at 100Hz, well above the maximum level of impedance to be measured.

As with the transfer function method, measurements were controlled using a LSIC30 based system. Due to the necessity of providing for quasi-simultaneous transfer function/impedance measurements (described in section 5.4) the serial port system was unavailable and the two channels of the onboard converters were used for the impedance measurements. A maximal length pseudo-random sequence of length 4096 was used to provide the excitation to the driving loudspeaker. To avoid delays and startup transients the last 2045 samples of the sequence were prepended onto the sequence giving a total length of 6141 samples. A pre-filter was designed and applied to the excitation sequence to boost the response above 2kHz, where the coupling between the two diaphragms begins to roll off. In a typical measurement the excitation signal is applied to the driving loudspeaker and simultaneously the velocity and pressure signals are recorded into the two channels of the 'C30 board. The acoustic impedance is found by taking the ratio of the FFTs of the pressure and velocity signals. The sampling rate in the time domain was chosen as 16.348kHz to give a frequency axis resolution of almost exactly 4Hz in the frequency domain.

5.4 The measurement rig

A measurement rig was designed for the purpose of conducting quasi-simultaneous impedance/transfer function measurements. The idea of this was to study the inter-relationship between the wall load (as measured by the impedance device) and the transfer function for different speakers and a range of

articulation.

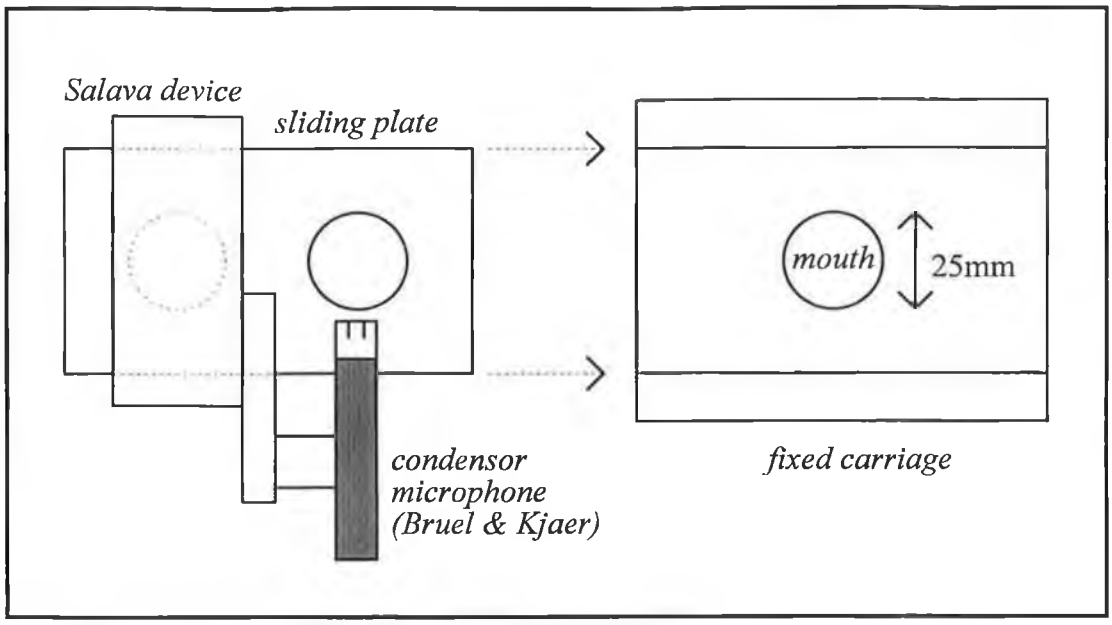


Figure 5.3 Front view of the mechanical slider system for joint measurements

The apparatus was designed to allow a rapid sequence of a transfer function measurement followed by an impedance measurement without the subject having to move or change articulation. The central part of the apparatus is the mechanical metal slider shown in figure 5.3.

The slider consists of two interlocking metal plates. One of the plates is kept fixed, and contains a 25mm diameter hole in the centre where the subjects lip-adaptor mouthpiece is inserted. The other is the sliding plate, inserted into the fixed carriage by means of grooves. The sliding plate contains two 25mm diameter holes and moves so that either of them is aligned with the hole in the fixed plate. An attachment for a condensor microphone is placed outside one of the holes in the sliding plate, and the impedance measurement head is placed in the other. The slider device was fitted on to one side of a 60cm by 60cm wooden baffle and aligned with a central 25mm diameter hole in the baffle. The purpose of the baffle was to shield the condensor microphone from direct radiation from the shaker in the transfer function measurements, and to provide a known plane radiating surface. With the centre of the baffle being at mouth level, a small glass window was added at eye level to allow the subject to respond to cues from a computer screen placed outside the studio window. For a dual

measurement a measurement protocol consisting of the following steps was developed and implemented in software.

1. The slider is placed in the transfer function measurement position, ie. the subjects mouthpiece is aligned with the condensor microphone.

2. A continuous noise sequence is played out to the shaker. The subjects seal their lips tightly around the mouthpiece, place the shaker to their neck listen through headphones to try and find the point that gives the loudest response.

3. The subjects hits a keyboard when ready and a transfer function measurement is taken.

4. The subject or an assistant moves the slider so that the impedance measurement head is aligned with the mouthpiece.

5. The subject or an assistant hits the keyboard and an impedance measurement is taken.

6. The results are calculated on the PC and displayed on the screen.

For steps 3, 4, and 5 above the subject must maintain the articulation. The duration of a transfer function measurement is 150ms and that of an impedance measurement 250ms so the task is not too difficult.

5.5 Processing the raw impedance data

A number of corrections have to be made to the raw impedance value measured by the Salava device. Firstly the impedance data has to be scaled to a reference level of 1cgs acoustic ohm ($gm.s^{-1}cm^{-4}$). The scaling factor was found by performing a calibration measurement on a stopped copper pipe of length 2m and diameter 25mm. The calibration is based on the principle that for an air column of fixed cross-sectional area near the measurement end, the impedance seen at that end becomes purely real at its poles and zeros, and that the local geometric mean of the poles and zero magnitudes is simply R_0 [Benade and Ibisi, 1987]. The tube length of 2m was chosen to give a high density of poles and zeros over the frequency range of interest. The required scaling factor (Zc) was derived from a dB scale plot of the measured impedance magnitude (shown in figure 5.4). The local geometric means of the poles and zeros were found

and a calibration curve was generated by joining these points. In practice the calibration curve was practically constant over frequency, so its average value was used.

As the calibration system is rigid and there is no movement during measurements, software was developed to perform a number of consecutive measurements, and average the results to improve noise performance. The impedance plot shown in fig. 5.4 is actually the average of a thousand consecutive measurements.

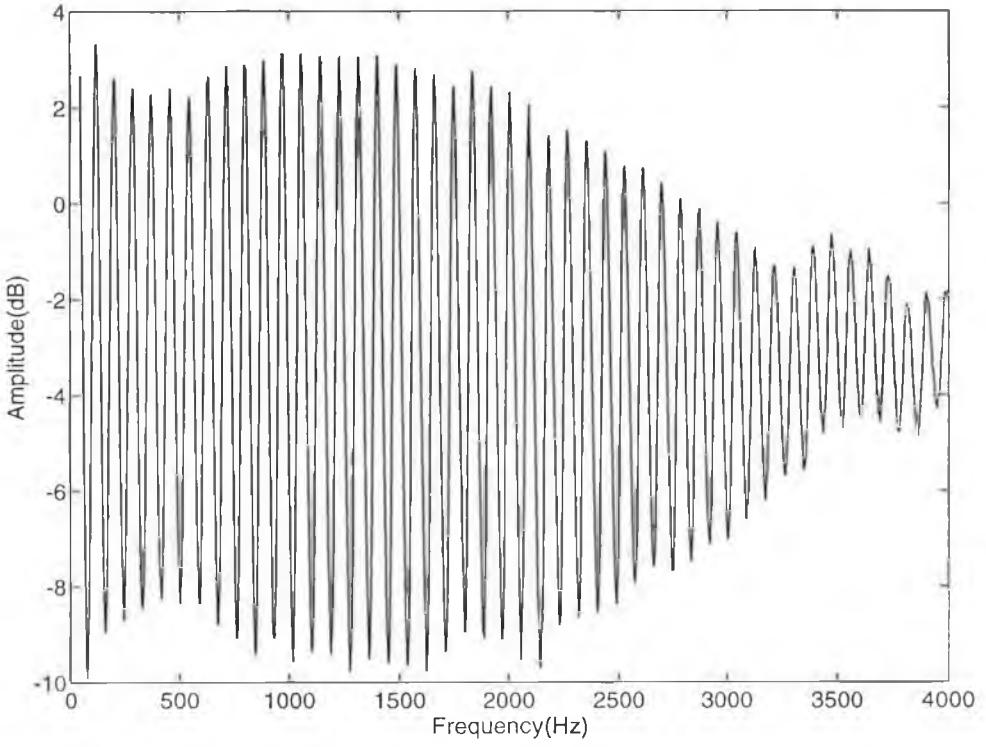


Figure 5.4: Measured acoustic impedance of a stopped 2m tube (unscaled)

The characteristic impedance of the measurement tube is given by the equation $Z_0 = \rho c/A$. In this equation A is the cross-sectional area of the measurement tube, ρ is the static density of air, and c is the speed of sound in air. To get from the raw data to a scaled value in acoustic-ohms we use.

$$\text{Zscaled} = \left(\text{Zmeasured}/Z_c\right) \times Z_0 \quad (\text{gm.s}^{-1}\text{cm}^{-4}) \qquad (5.4)$$

Another correction has to be made to account for the 8mm spacing between the microphone and the headphone diaphragm. If the pressure and the flow variables at the microphone are denoted P_m and U_m respectively, and those at the headphone diaphragm P_h and U_h, then:

$$Z_{hh} = P_h / U_h \qquad \text{and} \qquad Z_{mh} = P_m / U_h \tag{5.5}$$

Z_{mh} is the measured and scaled impedance, and Z_{hh} is the desired impedance seen at the walkman headphone. The relationship between these quantities is given by Keefe and Benade [1981].

$$Z_{hh} = (Z_{mh} + j.Z_0 \sin(ke))/\cos(ke) \tag{5.6}$$

where $k = \omega/c$ and e is the microphone diaphragm spacing in cm. This correction is automatically applied when an impedance function is being generated from raw pressure and velocity signals.

A further correction must also be made to the measurements on human subjects to take account of the effect of the lip adaptor tube used in the measurements. This correction is described in appendix C. Figure 5.5 below shows a wideband lip impedance measurement made using the apparatus.

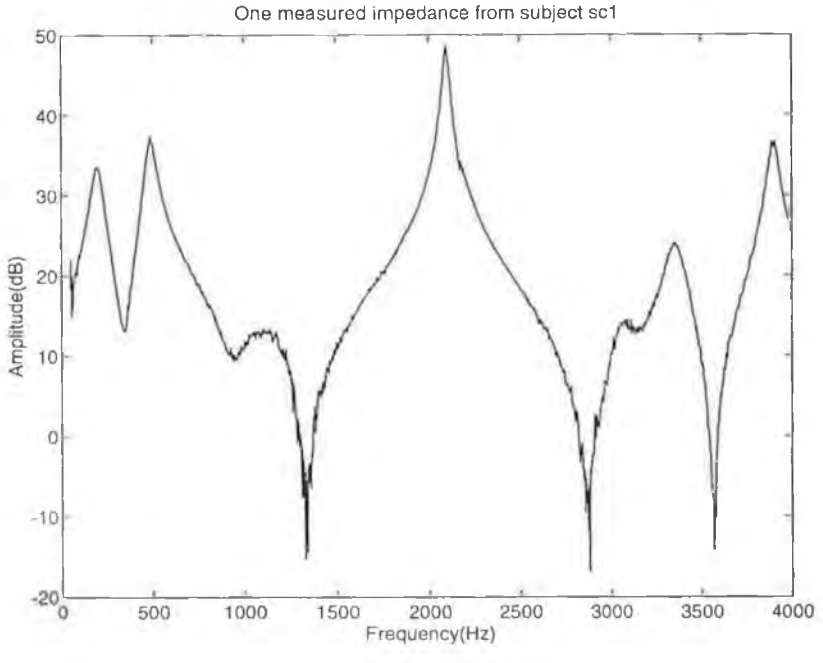


Figure 5.5 Measured wideband lip impedance, subject sc1

5.6 The Fant, Branderud and Nord model

In human speech the first formant is known to decrease with decreasing lip opening. The extreme lower limit of F1 occurs when the lips are shut, and is known as the

closed tract resonance. This resonance corresponds to the first resonance of the vocal tract acoustic input impedance seen at the lips. According to [Fant et al, 1976], the low frequency equivalent circuit of the closed vocal tract is simply a capacitor $C_t = V_t/\rho c^2$ in parallel with an inductor in series with a resistor. The inductor, L_w, represents the lumped mass element of the vocal tract walls, and the resistor, R_w, determines the damping of the system. The equivalent circuit is shown in figure 5.6 below. A capacitor, C_w, in series with L_w is also shown. C_w represents the compliance of the vocal tract walls. Fant el al have estimated that C_w provides a series resonance with L_w at about 30Hz-70Hz, which is well below the range of the closed tract resonance and accordingly less important. If C_w is neglected the input impedance of the low frequency model has one resonance corresponding to the closed tract resonance. The question now arises, how do we extract the equivalent circuit parameter values from a measurement of this resonance. The measurement system used by Fant et al was not absolutely calibrated. This meant that they had to insert an extra lip calibration tube into the subject's mouth during the measurements. One measurement was made with the calibration tube open, and a second measurement was made with it shut. The known acoustic inductance of the calibration tube was added to the equivalent circuit model, and from an analysis of the two measurements they could calculate the circuit component values. In our case the system was absolutely calibrated so a second measurement was not necessary.

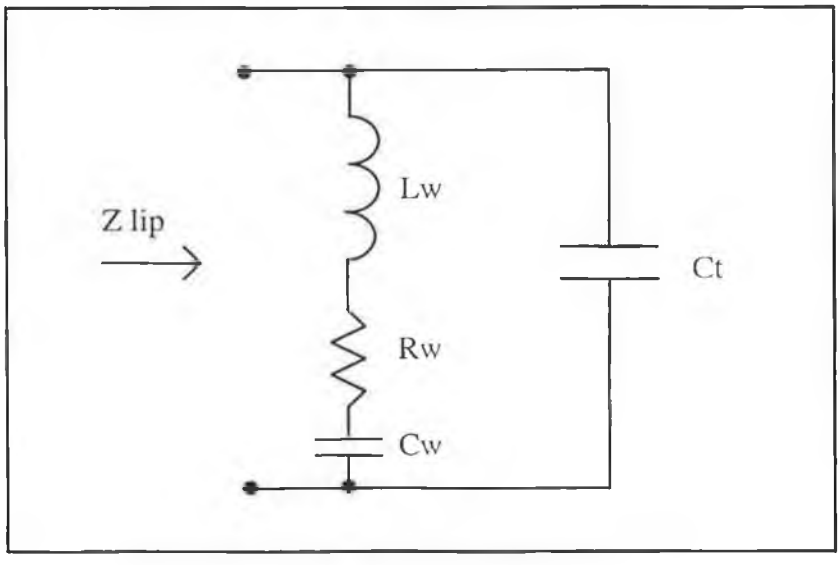


Figure 5.6: Low-frequency equivalent circuit of the vocal tract

Software was developed to automatically extract the calibrated values of peak amplitude ($|Z|_{max}$), peak frequency ($f_m = \omega_m/2\pi$), and bandwidth (β) from a measurement of the closed tract resonance. The input impedance Z_{lip} of the equivalent circuit above is (note C_w is neglected):

$$Z_{lip} = \frac{R_w + j\omega L_w}{1 - \omega^2 L_w C_t + j\omega R_w C_t} \tag{5.7}$$

The resonance frequency (ω_0) of this circuit is defined as the frequency that makes this impedance purely resistive. In general, ω_0 does not occur at the same frequency as ω_m, the frequency of the maximum value of $|Z|$, and this makes the analysis of the circuit complicated. Analysis shows that:

$$\omega_0 = \sqrt{\frac{1}{L_w C_t} - \left(\frac{R_w}{L_t}\right)^2} \tag{5.8}$$

ω_m is found by differentiating $|Z_{lip}|$ with respect to ω, and finding the value of ω that makes this expression equal to zero. We find that:

$$\omega_m = \left(\sqrt{\left(\frac{1}{L_w C_t}\right)^2 - \left(\frac{R_w}{L_w}\right)^2 \left(\frac{2}{L_w C_t}\right)} - \left(\frac{R_w}{L_w}\right)^2\right)^{\frac{1}{2}} \tag{5.9}$$

The Q factor of a resonance is defined as the ratio of the resonance frequency to the bandwidth ($Q_0 = \omega_0/\beta$). If Q_0 is high analysis shows that $\beta = R_w/L_w$. We see from the above equations that as R_w/L_w goes to zero, then $\omega_0 \equiv \sqrt{1/L_w C_t}$ and $\omega_m \equiv \omega_0$. In this case we get $Q_r \equiv Q_0$ where $Q_r = \omega_m/\beta$. Further analysis shows that we can arrive at the following expression for the peak impedance magnitude:

$$|Z|_{max} = Q_r^2 . R_w \tag{5.10}$$

Algebraic manipulation of these equations gives the following simple expressions for the desired equivalent circuit component values (R_w, L_w, C_t) in terms of the measured resonance parameters ($|Z|_{max}$, f_m, β).

$$R_w = \frac{|Z|_{max} . \beta^2}{f_m^2} \quad (5.11a), \quad L_w = \frac{R_w}{\beta . 2\pi} \quad (5.11b), \quad C_t = \frac{\beta}{2\pi . R_w . f_m^2} \quad (5.11c)$$

Note that the derivation of these simple expressions is based on the assumption of a high Q factor. Textbooks differ on what is meant by a high Q factor. To check the

validity of this assumption closed tract resonances were simulated using equation 5.7 and values of R_w, L_w, and C_t obtained from Fant's data. The parameters $|Z|_{max}$, f_m, and β were extracted from the simulated resonances, and used to calculate estimates for R_w, L_w, and C_t using equations 5.11 above. These were then compared to the original values. In general, the calculated values of R_w, L_w, and C_t differed from the originals by about 20%. This proved that the assumption of a high Q was not valid, and the resulting simplifications gave erroneous results.

Without this simplification the expressions for R_w, L_w, and C_t in terms of the measured parameters $|Z|_{max}$, f_m, and β become highly non-linear and extremely difficult if not impossible to solve analytically. Instead of tackling these equations a numerical solution to the problem was sought. Firstly a polynomial was fitted to the measured resonance to get it as a mathematical expression ($y(\omega)$). A 7th order polynomial was found to give a good fit. Next an iterative scheme was applied to minimise the least-squares difference between the experimental curve and the model curve, where the model curve is:

$$|Z(\omega)| = 20 \times \log_{10}\left(\frac{R_w^2 + \omega^2 L_w^2}{(1-\omega^2 L_w C_t)^2 + \omega^2 R_w^2 C_t^2}\right)^{\frac{1}{2}} \qquad (5.12)$$

and the function to be minimised is:

$$F(R_w, L_w, C_t) = \sum_{\omega}\left[(y(\omega) - |Z(\omega)|)^2\right] \qquad (5.13)$$

The minimisation was performed using the MATLAB optimisation toolbox. The algorithm needs a starting guess to the values of R_w, L_w, and C_t. It then iterates until it converges to the values of R_w, L_w, and C_t that minimise $F(R_w, L_w, C_t)$ to a specified tolerance. The speed of convergence is greatly increased if the approximate equations 5.11 are used to calculate starting guesses for R_w, L_w, and C_t. The method was tested in the same way as before. Resonances were simulated with known values of R_w, L_w, and C_t. Values of $|Z|_{max}$, f_m, and β were extracted from the resonance and used in equations 5.11 to give starting guesses for the minimisation. A polynomial was fitted to the resonance and the minimisation performed. The resulting extracted estimates of R_w, L_w, and C_t corresponded exactly to the originals. The method was then used to extract component values from measured resonances. The resonances

were then resimulated from the extracted values to check the fit. In all cases the fit was very good. An example is shown in figure 5.7 below.

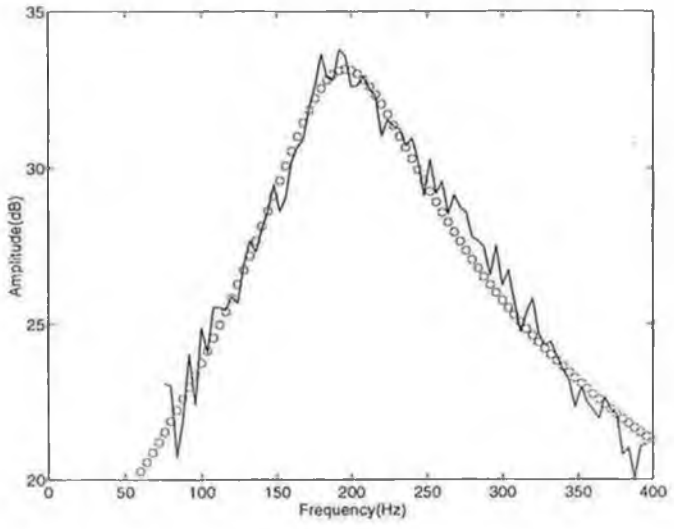


Figure 5.7: Measured lip impedance (solid), and simulated impedance (circles) from the extracted circuit component values

5.7 Summary:

In this chapter the theory and practice of vocal tract input impedance measurements was discussed. The measurement rig set up to perform quasi-simultaneous vocal tract acoustic transfer function and vocal tract acoustic input impedance measurements was discussed. The post measurement processing needed to transform the measured results into the desired input impedance over frequency form and to extract wall vibration load parameters was described. In the next chapter a survey of the lip-impedance measurements on a wide range of human subjects is described.

Chapter 6

The Wall Vibration Load Survey

6.1 Introduction

A survey of the wall vibration load was carried out on human subjects using the Salava device and the measurement rig described in section 5.4. A wide range of subjects were used, children, adults, male and female. The subjects were instructed to maintain an [a] articulation keeping the glottis closed to block off the subglottal cavities, and the velum raised to avoid shunting in the nasal cavities. Most of the subjects were not phonetically trained so the general procedure was to perform a few trial measurements; displaying the results on the PC screen each time until the subjects became comfortable with the apparatus, and felt that they could reproduce the same articulation.

For an impedance measurement the subject closes his/her lips tightly around the lip adapter tube, forms an [a], and signals the assistant when ready. For each subject a batch of 12 measurements were taken. In total measurements were performed on 33 subjects. Of these 24 produced useful results. The remainder of the subjects seemed to have a problem maintaining a consistent articulation without moving or breathing, and in some cases did not properly seal their lips around the adaptor. For each of the 24 useful subjects the results were further pruned to remove obvious outliers, which again could occur for the reasons given above. For each subject the raw data was processed as decribed in appendix C to scale it and correct for the lip adaptor tube. The data was then further processed as described in section 5.6 to extract the low-frequency wall vibration information.

6.2 The Survey results

The results of the survey showed two distinct populations. For about half of the subjects a single broad resonance was observed in the low-frequency region of the measurement, with a peak frequency of about 200Hz. For the other half two narrower resonances were observed in the same frequency region. Typical examples of both types are shown in figure 6.1 below.

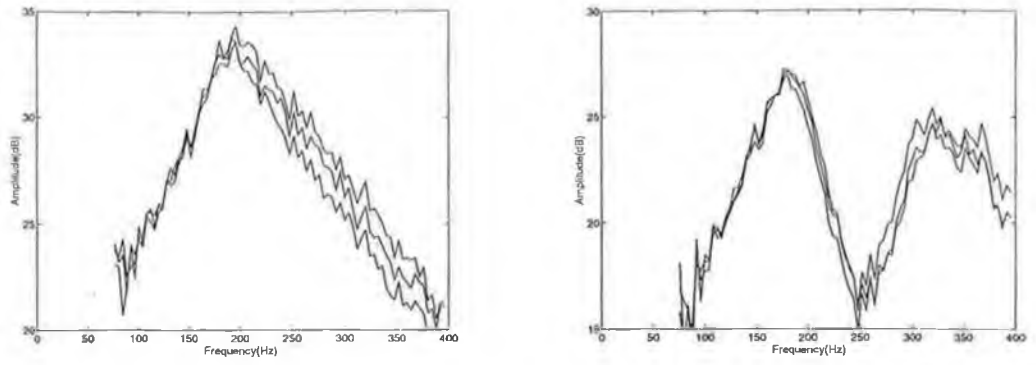


Figure 6.1: Typical examples of 'single peak' and 'double peak' subjects. The left shows three measurements on subject dc2, the right shows three measurements on subject rn1.

Subject Code	Subject type	Mean Frequency (Hz)	Mean Bandwidth (Hz)	Mean Amplitude ($gm.s^{-1}cm^{-4}$)
cl1	adult female	199.9	110.3	28.84
gr1	adult male	241.1	113.7	36.58
mk1	male 10 years	259.3	75.7	32.41
hl1	female 4 yrs	287.8	143.9	38.18
rh1	female 4 yrs	299.9	197.2	38.36
pl1	adult female	223.3	145.9	34.92
pc1	adult female	200.4	110.9	33.22
sc1	adult male	243.4	77.5	31.09
dv1	adult male	192.9	91.7	34.02
dc2	adult male	199.9	95.7	33.51
yi1	adult male	179.4	87.0	28.91
nl1	adult male	217.5	116.8	33.28
bh3	adult male	189.2	104.0	26.82

Table 6.1 : Mean closed tract resonance parameters for 'single peak' subjects.

The second peak cannot be explained by our closed tract model. The question arises, is this peak due to nasal or subglottal coupling, or is it possibly even a natural effect with particular subjects. Our hypothesis is that the second peak occurs due to nasal coupling and is discussed in section 6.4. Results extracted for the single peak subjects are tabulated above in table 6.1. The standard deviations for these measurements showing the intra-subject consistency are given below.

Subject Code	Subject type	Standard Deviation Frequency (Hz)	Standard Deviation Bandwidth (Hz)	Standard Deviation Amplitude ($gm.s^{-1}cm^{-4}$)
cl1	adult female	16.9	3.5	0.78
gr1	adult male	20.6	10.2	0.80
mk1	male 10 years	2.7	3.2	0.44
hl1	female 4 yrs	3.9	0	0.08
rh1	female 4 yrs	<5 samples	<5 samples	<5 samples
pl1	adult female	9.2	10.9	1.04
pc1	adult female	5.8	9.5	0.17
sc1	adult male	3.3	2.0	0.15
dv1	adult male	5.0	10.3	0.53
dc2	adult male	6.0	9.5	0.32
yi1	adult male	3.3	6.6	0.51
nl1	adult male	18.2	8.6	0.73
bh3	adult male	9.2	14.4	0.53

Table 6.2: Standard deviations within a batch of up to 12 measurement for the parameters given in table 6.1

It can be seen from table 6.2 above that most of the subjects are reasonably consistent over the batch of measurements. Table 6.3 below shows the mean equivalent circuit component values extracted for each of these subjects.

Subject Code	Subject type	Equivalent Resistance R_w $gcm^{-4}s^{-1}$	Equivalent Inductance L_w gcm^{-4}	Equivalent Capacitance C_w $g^{-1}cm^2$ $1\times10^{-4}\times$	Vocal Tract Volume cm^3
cl1	adult female	5.86	0.0095	0.644	91.8
gr1	adult male	11.64	0.0182	0.249	35.4
mk1	male 10 yrs	4.08	0.0078	0.505	71.2
hl1	female 4 yrs	15.39	0.0188	0.162	23.2
rh1	female 4 yrs	23.38	0.0206	0.126	18.1
pl1	adult female	16.60	0.0196	0.239	34.2
pc1	adult female	10.09	0.0150	0.366	52.2
sc1	adult male	3.06	0.0071	0.685	97.7
dv1	adult male	9.25	0.0168	0.397	56.6
dc2	adult male	8.88	0.0152	0.399	57.0
yi1	adult male	5.19	0.0104	0.814	116.0
nl1	adult male	9.22	0.0143	0.361	51.5
bh3	adult male	4.58	0.0078	0.847	120.7

Table 6.3: Equivalent circuit component values extracted for the 'single peak' subjects

The calculated volumes vary widely, but there are some underlying trends. Fant et al measured an average vocal tract volume of about $98cm^3$ for the average male, and about $53cm^3$ for the average female. These results agreed with an earlier study by Fant [1972] on x-rays of Russian vowels. However their impedance measurements also showed inter-subject standard deviations for both of these results of $39cm^3$ and $20cm^3$ respectively. This seems to suggest that much depends on the subject's actual tract configuration during the measurement. Both of the four year old girls measured show volumes of about $20cm^3$, and one of them (hl1) was our most consistent subject. For the double peak subjects, the parameters of the first peak were extracted, and are tabulated in table 6.4 below.

Subject Code	Subject type	Mean Frequency (Hz)	Mean Bandwidth (Hz)	Mean Amplitude ($gm.s^{-1}cm^{-4}$)
rn1	adult male	181.3	62.2	26.78
bh1	adult male	187.9	63.9	24.87
ks1	adult female	222.3	92.7	23.44
ab1	adult female	212.7	76.0	25.68
es1	adult female	197.9	72.0	26.23
yl1	adult female	197.1	58.3	28.28
cl2	adult female	182.9	56.9	26.34
ad1	adult male	157.9	69.6	23.47
kn1	adult male	210.5	57.3	29.07

Table 6.4: Extracted parameters of the first peak for the 'double peak' subjects

The bandwidth of the first peak for the double peak subjects is on average about half of the bandwidth of the peak for the single peak subjects (67Hz vs 113Hz). Our closed tract model can also be fitted to this peak. However, the calculated values for the equivalent circuit components, and the resulting VT volumes suggest that in this case the assumption of a closed vocal tract, on which our model is based, is not valid. The possible causes of the second peak are discussed in section 6.4. Here we compare the results for the 'single peak' subjects with the Fant study.

6.3 Comparison of the 'single peak' results with the Fant study

If we assume our single peak subjects had their velums raised, we can compare the results with the Fant [1976] study. The table shown below presents the results (HKS denotes Hogan, Krietmeyr, Scaife). Some differences should be pointed out. In the Fant study, an attempt was made to separate the wall load component concentrated near the lips from that concentrated in the pharynx region; the Fant values for R_w are their estimates for the pharyngeal region only. Our measurements show the aggregate V_t, R_w and L_w obtained for the entire vocal tract with a neutral articulation.

	F_w (Hz)	B_w (Hz)	L_w $(g.cm^{-4})$	V_t (cm^{-3})	R_w $(g.s^{-1}cm^{-4})$
5x M Adults (Fant)	191	76	0.011	98	18
inter-subj. std. dev.	(15)	(4)	(0.004)	(39)	(--)
5x F Adults (Fant)	218	94	0.014	53	15
inter-subj. std. dev.	(27)	(11)	(0.006)	(20)	(--)
7x M adults (HKS)	209	98	0.013	76	7
meas. std. dev.	(8.7)	(9.5)	(0.0006)	(5.5)	(0.8)
inter-subj. std. dev.	(25.4)	(14.2)	(0.0044)	(34)	(3.1)
3x F adults (HKS)	208	122	0.015	59	11
meas. std. dev.	(12)	(9)	(0.0013)	(6.3)	(1.5)
inter-subj. std. dev.	(13.3)	(20.4)	(0.0051)	(29.5)	(5.4)
1x M 10 yr. old (HKS)	259	76	0.008	72	4
inter-subj. std. dev.	(n/a)	(n/a)	(n/a)	(n/a)	(n/a)
2x F 4yr. old (HKS)	294	171	0.020	21	19
inter-subj. std. dev.	(n/a)	(n/a)	(n/a)	(n/a)	(n/a)

Table 6.5: Averages and standard deviations for measurements of wall load parameters Fant et al 1976 and the present study (HKS) .

In the table there are two standard deviations quoted. The measurement standard deviation is the r.m.s average of the standard deviations for each subject in the group. and is a measure of the intra-subject measurement variation. The inter-subject standard deviation is the r.m.s average of the subject averages, and is a measure of the spread of data between subjects. As can be seen from the table, the inter-subject s.d. is in most cases much larger than the measurement s.d., indictating that inter-subject differences are significant.

Because the sample size is small we must be wary about generalising from the results. The results for the 4yr old children show, as expected a much lower volume than the adults, and a substantially greater acoustic inductance term. Subject MK1, a 10 yr old

boy, is intermediate in location between the adult and infant groups. For the present study there is little difference in the resonance centre frequency between the male and female subjects unlike the Fant study. However the cavity bandwidth terms are significantly greater for the female subjects in both studies. The mean wall inductance terms are very similar for both sets of measurements, e.g. for adult male subjects $0.0013 g.cm^{-4}$ in the case of the HKS study versus $0.011 g.cm^{-4}$ in the Fant study. The inter-subject standard deviations are also very similar here, $0.0044 g.cm^{-4}$ or the HKS study versus $0.004 g.cm^{-4}$ for the Fant study. This could be an indication that a wide spread of values is to be expected in measurements of this type.

6.4 Effects of glottal and nasal coupling on the vocal tract input impedance

An extra resonance with a peak frequency of about 300Hz (and not explained by our low-frequency model) is observed in about 50% of the subjects. Our hypothesis is that the second peak occurs due to nasal coupling. The main reason for this is that the velum is much harder to control than the glottis during the measurements. The subjects are instructed to close their glottis by holding their breath. Thus there is no breathing during the measurement, and it becomes correspondingly more difficult for the subject to know whether or not the velum is fully raised. Another reason to reject the alternative hypothesis that the peak is due to subglottal coupling is that according to data on human lung impedance measurements published by Ishizaka et al [1976], the first resonance peak of the subglottal system should occur at about 600Hz, however the second peak in all the subjects we measure seems to occur consistently at about 300Hz (Note: This agrees with the results of earlier measurments on the cadavers of large dogs by Van den Berg [1958], However Ishizaka's measurement were conducted on live Humans (laryngectomized subjects))

To test our hypothesis we adopted a dual approach.

1) Measurements on a trained speaker.

2) Simulation of glottal and nasal effects.

6.4.1 Measurements on a trained speaker

Our subject was a trained phonetician, and had a knowledge of the articulations necessary (within the constraints of the experiment) to achieve the four test vocal tract

configurations. The four test configurations were as follows (in all cases the articulation was a near to an [a] as possible):

a) Closed glottis, closed velum.

b) Open glottis , open velum.

c) Open glottis, closed velum.

d) Closed glottis, open velum.

The results were as follows:

a) A single broad peak with a maximum at around 200Hz.

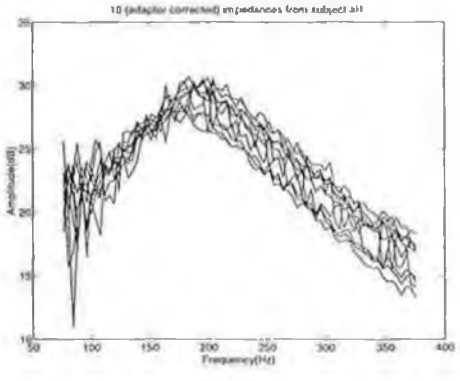
b) A broad peak with a maximum at just over 200Hz, and a second peak appearing in the vicinity of 400Hz

c) A single broad peak, with a maximum frequency atapprox 240Hz.

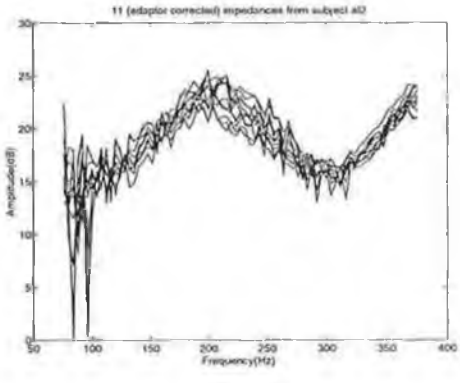
d) Two Narrower peaks with the first maximum at around 180Hz, and the second at around 320Hz.

These results are shown below in figure 6.2.

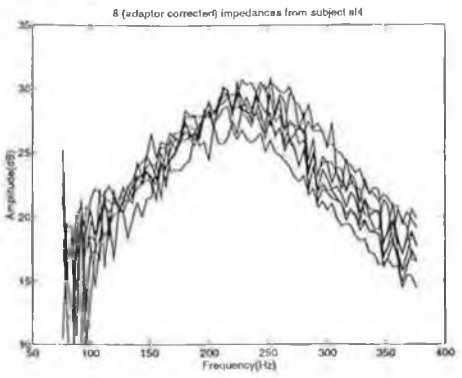
Closed Glottis, Closed Velum



Open Glottis, Open Velum



Open Glottis, Closed Velum



Closed Glottis, Open Velum

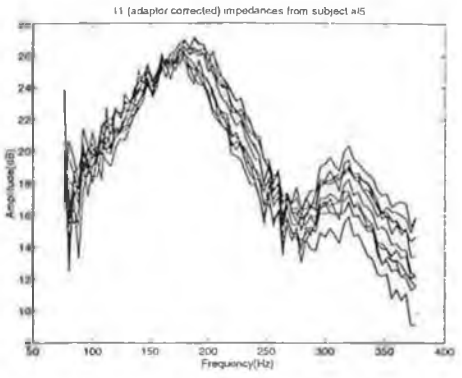


Fig 6.2: Measurements of a trained subject [Al1], for four different articulations.

These results support the hypothesis. A double peak only appears when there is nasal coupling. Also, there is the result that without nasal coupling, glottal coupling has the effect of raising the frequency of the first input impedance resonance.

6.4.2 Effect of glottal coupling on the vocal tract input impedance.

To simulate the effect of glottal coupling on the first resonance of the input impedance we need an equivalent circuit that takes into account the subglottal system, the area dependent glottal impedance, and the input impedance of the supraglottal system. The subglottal equivalent circuit and component values used was taken from Badin and Fant [1984] and consists of three parallel *rlc* circuits (see figure 6.3).

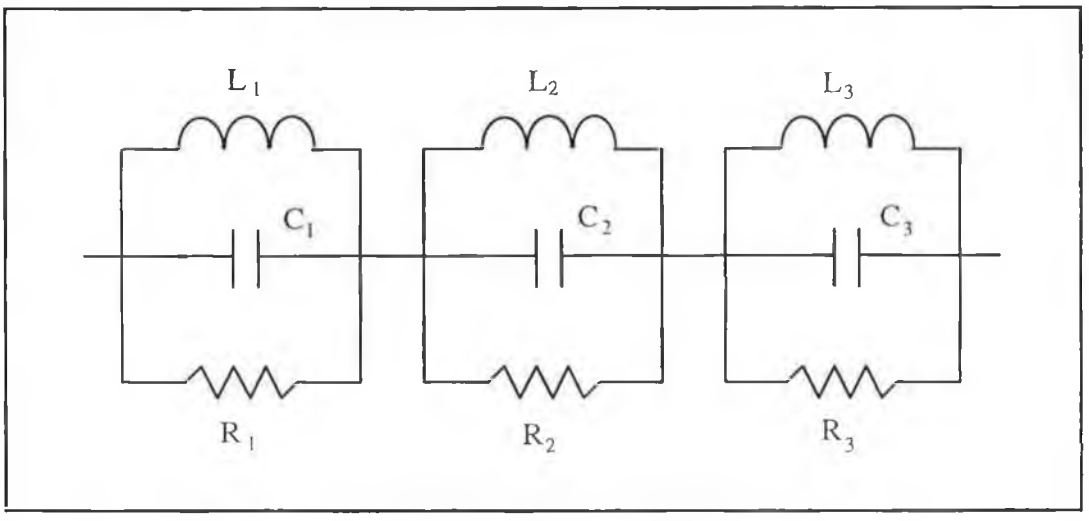


Figure 6.3: Equivalent circuit of the subglottal system

A plot of the magnitude of the input impedance of the subglottal model is shown in figure 6.4 below. The area dependent glottal impedance model was taken from Fant [5], and consists of a resistor and inductor in series. In this model:

$$R_g = \frac{0.17 \times 0.64 \times \rho c}{A_g} \quad \text{(6.1a), and} \quad L_g = \frac{\rho \times 0.5}{A_g} \quad \text{(6.1b)}$$

where $\rho = 0.00121 g.cm^{-3}$, $c = 34318 cm.s^{-1}$ and A_g is the glottal cross-sectional area in cm^2. The glottal impedance is thus:

$$Z_g = R_g + j\omega L_g \quad \text{units: } g.cm^{-4}.s^{-1} \text{ (acoustical ohms).} \quad (6.2)$$

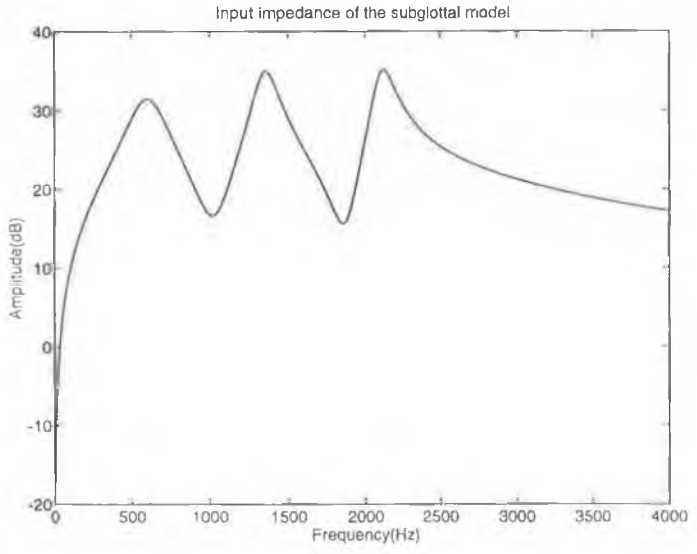


Figure 6.4: Input impedance of the subglottal model

The supraglottal model used was that of Fant, Nord and Branderud [1976] and consists of a capacitor in parallel with a resistor and inductor in series. In this simple model the capacitor is analogous to the volume of air in the vocal tract, the inductor is analogous to the lumped mass element of the vocal tract walls, and the resistor determines the energy losses. The component values for this circuit were taken from actual measured values on subject [cl1] (one of the single peak subjects). A file was written to compute and plot the input impedance of the complete circuit for different values of *Ag*, the glottal cross-sectional area.

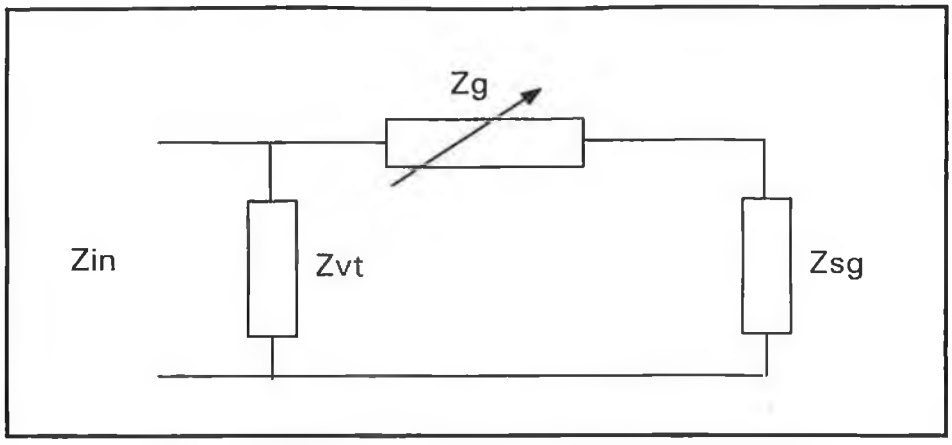
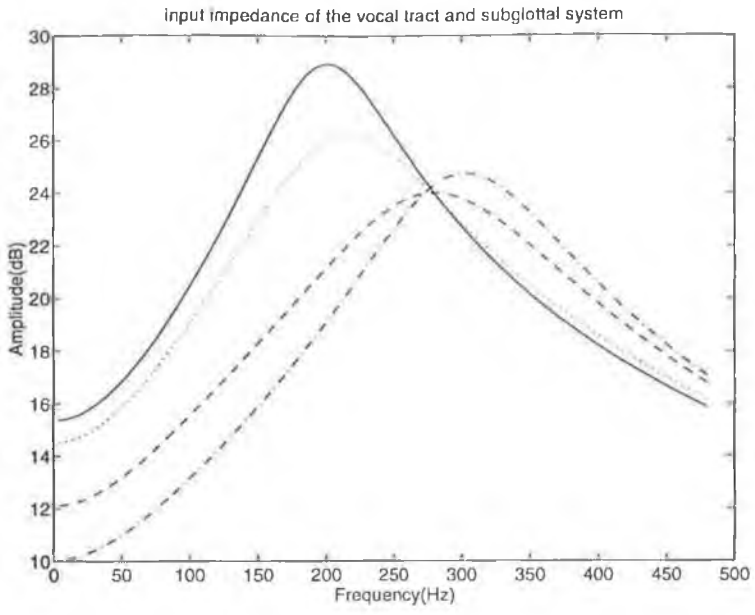


Figure 6.5: Equivalent circuit of the vocal tract and subglottal system used in the simulation.

where:
$$Z_{in} = \frac{Z_{vt}(Z_g + Z_{sg})}{Z_{vt} + Z_g + Z_{sg}} \tag{6.3}$$

According to Fant [1972] the glottal area ranges from $0cm^2$ (fully closed) to $0.65cm^2$ fully open. In voicing the glottal area only reaches $0.08cm^2$ at maximum opening. With a glottal area of $0cm^2$, the glottal impedance becomes infinite and the input impedance we observe is just the input impedance of the supraglottal circuit. As the glottal area increases the glottal impedance decreases and the coupling to the subglottal system becomes greater. The simulation showed that increasing the glottal area had the effect of broadening the first peak of the VT input impedance, and raising the frequency. This agrees with our measured results on the trained speaker above.



Solid line (-): ag=0cm^2 Dotted line (...): ag=0.08cm^2

Dashed line (--): ag=0.35cm^2 Dashdot line (-.): ag=0.65cm^2

Figure 6.6: Variation of the vocal tract input impedance with varying glottal area.

6.4.3 Effect of nasal coupling on the vocal tract input impedance.

According to Flanagan [1972], in the adult male the nasal cavity has a length of about 12cm and a volume of the order of 60cm^3. The degree of coupling between the vocal tract and the nasal cavity is determined by the cross-sectional area of the velar opening, and this in turn is controlled by the velum, a thin muscle that can be raised or lowered to allow varying amounts of nasal coupling. As the velum is lowered increasing coupling between the nasal tract and the vocal tract, it moves towards the tongue. This causes a constriction as seen from the mouth end, and changes the whole vocal tract configuration. Flanagan states that the area of the velar opening can range from 0cm^2 to 5cm^2. Fant [1972] describes simulations of the nasal tract using an electrical line analog (LEA). A number of unit filter sections, each simulating an acoustic section of effective length 1cm and a variable area, were combined in cascade to form a nasal branch of the electrical analog. His simulations show that for a velar coupling area of 0.16cm the first peak of the input impedance of the nasal tract as seen from the uvula

from the uvula occurs at 300Hz., which is the same frequency of the second peak seen in our 'double peak' subjects.

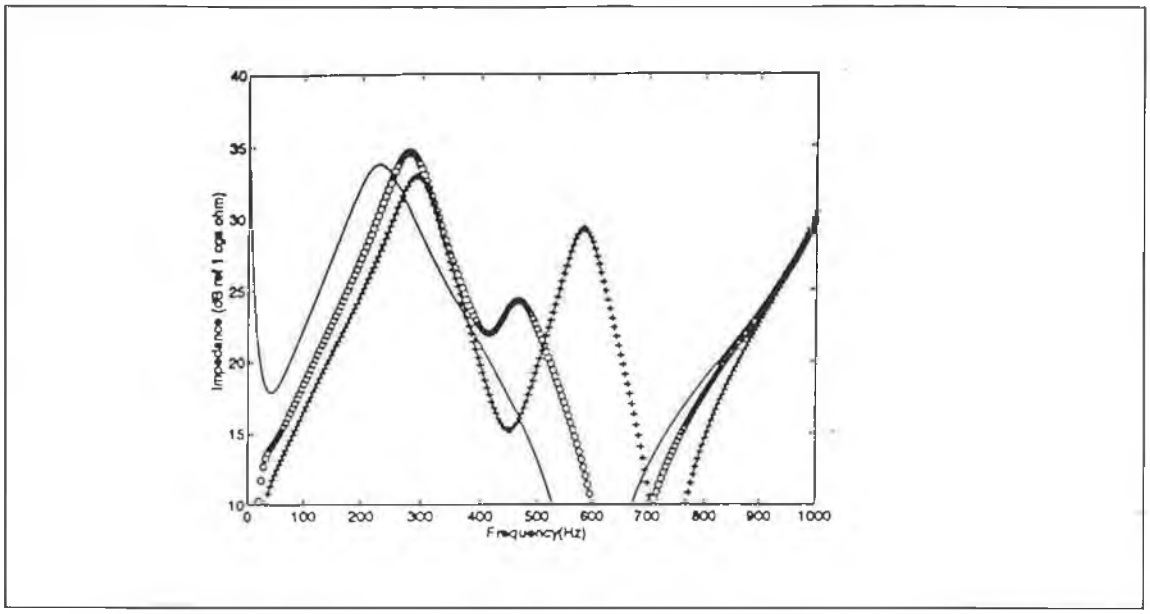


Figure 6.7: Simulated lip impedance for an [a] articulation, using:

----- no nasal coupling

oooo 0.1cm^2 nasal coupling area

+++ 0.3cm^2 nasal coupling area.

The degree of 'sharpness' of this peak depended on the value of nose cavity damping resistor used in the simulation. He noted that this peak increases in frequency as the coupling area is increased. Simulations were made in DCU with varying degrees of nasal coupling. For this purpose a modified version of the Maeda vocal tract simulator was used (see section 5.6). The nasal tract model in the simulator is a simple transmission line with losses. It is inserted into the oro-pharyngeal tract model by means of a variable coupling area. The simulations showed that as the velar coupling area was increased, secondary peaks appeared in the lip impedance function, see figure

6.7. The centre frequency of the peaks appear at much higher centre frequencies than the experimentally observed values, but it should be noted that the model of the nasal tract in the simulator is relatively simple. There is one channel 13cm long rather than two seperated by a septum. Also, there is no model of the sinuses. It is possible that the secondary peak observed at 300Hz might be better modelled in a more sophisticated nasal tract model.

6.5 Summary

In this chapter the acoustic impedance measurements carried out at DCU have been described. A survey of lip-impedance load data was carried out to determine the spread of parameter values for a lumped wall impedance model. A wide range of speaker types were used, including male, female and children. It was noted that the results could be broken down into 'single peak' and 'twin peak' subjects and reasons to explain this effect were proposed and tested. Further experiments on larger sample sizes are needed to confirm these results. The results were compared with the Fant study on vocal tract lip impedance. Similar results were obtained for the wall inductance parameter of the closed tract equivalent circuit. In the next chapter we summarise the results of these acoustic measurements, and discuss how these results are now being used to assist in the study of articulatory modelling.

Chapter 7

Conclusions

7.1 Summary and statement of conclusions

The main aim of this thesis has been to contribute to the modelling of human speech production using data gained from acoustic measurements of the human vocal tract. It is agreed by speech researchers that more real data is required, especially when adapting a reference model to a specific speaker. The continuing challenge of speech modelling is summarised in [Fant 1980]: ".. to improve techniques for inferring vocal tract characteristics from speech wave data we need a better insight into vocal tract anatomy, area function constraints, and a continued experience of confronting models with reality. A balanced mixture of academic sophistication and pragmatic modelling". With this vision in mind, DSP based measurement systems were set up, and acoustic impedance and transfer function measurements were performed on a wide variety of subjects in DCU's semi-anechoic chamber. The measurement systems are compact, and can be reproduced by researchers in other laboratories.

The DSP based systems use efficient algorithms to process and display the data, allowing the subject and the experimenter to view the results quickly. Thus, feedback is enhanced allowing the subject to try many different configurations, and achieve a 'feel' for articulatory-acoustic relations. For example, using direct feedback of results, we can train so called 'twin peak' (see chapter 6) subjects to reduce coupling to the nasal tract by raising their velum. Quantitavely, the experimental methods allow us to extract wall vibration load parameters, formant frequencies, formant bandwidths, and lip impedance peaks for different subjects.

The measurement rig has been used in a survey of lip impedance data to determine the spread of parameter values for a lumped impedance model, including male and female adult subjects and some children. The results point to the possible need in acoustic to articulatory inversion, of adapting the reference model to specific subjects by separately estimating the wall impedance load.

The vocal tract wall vibrations occuring along the vocal tract have their greatest amplitude in a region about 4cm from the glottis [Badin and Fant, 1984]. These

effects were included at the glottal end of the Maeda speech production model [Maeda, 1982], and used to estimate the effect of measured wall vibration parameters on formant frequencies.

7.2 Future developments

It is hoped that the measurement systems developed during this work will be further enhanced and developed by others. The facility is now available to conduct large scale experiments on human vocal tract lip-impedance, and supplement the results with transfer function data. Further experiments on children would be particularly useful, as this sort of data is practically non-existent and yet necessary for adaptable models.

The transfer function apparatus could be modified without too much difficulty (eg. using miniature nostril microphones as in recent experiments (unpublished) at ICP Grenoble) to measure nasal tract transfer functions. Data on nasalised vowels is scarce, and existing models of the nasal tract are still considered primitive.

An articulatory codebook is a list of ordered pairs of vectors. One vector of the pair describes a geometrical shape of the vocal tract; the other describes the acoustic properties of the vocal tract shape. The usual way of employing such a codebook is to retrieve the vocal tract shape that corresponds to the acoustical parameters that best match the spectrum of a given frame of speech [Schroeter and Sondhi, 1989]. A simple codebook containing 960 vocal tract shapes has already been developed at DCU [Jeevakumar, 1993]. Further work in this area will include investigation of adapting codebooks to specific speakers using parameters derived from acoustic measurements.

Further analysis of the results obtained from the wall vibration load survey (see chapter 6), is now underway in DCU. Preliminary studies and simulations have been undertaken in which the measured values of Lw and Rw have been used [ESPRIT/BR No 6975 Deliverable 29]. The object of the simulations was to find, for a perturbation of Lw and Rw within the measured range of the survey results, the magnitude of adjustments in the seven shape parameters of the Maeda [Maeda 1989] articulatory model needed to optimally restore the original values of F1, F2, F3 and B1 for a number of standard vowels. The initial simulation results are summarised as follows. "In summary, for the range of vowel shapes considered here, the compensatory

adjustments of articulatory parameters due to incorrect assumptions about the wall vibration load, within observed limits of variation, are quite small, of the order of 10-20% of one standard deviation". Further simulations using the Maeda model and the measured wall vibration parameters are yet to be carried out.

References

Badin, P. and Fant G. (1984): "Notes on vocal tract computation", STL-QPSR 2-3/1984. pp 53-108.

Badin, P. et al (1993), "Acoustic transfer functions for vowels and consonants", Speech Maps (ESPRIT/BR 6975), Deliverable 4. Appendix D.

Benade, A.H. and Ibisi, M.I. (1987). "Survey of Impedance Methods and a new Piezo-Disk-Driven Impedance Head for Air Columns", J. Acoust. Soc. Am., 81(4), pp 1152-1167.

Burke, P. (1989): "Acoustic Measurement of the Area Function of the Vocal Tract", BE project thesis, Dept. of Electronic Engineering, University College Galway, May 1989 (unpublished).

Castelli, E. (1989): "Caraterisation Acoustique des Voyelles Nasales du Francais", Docteur Thesis, INPG Grenoble.

Coker, C.H. and Fujumura O. (1966). "A model for specification of vocal area function", J.Acous. Soc. Am. 90(5), pp. 2394-2409.

Devaney, J.W. and Goodyear, C.C. (1994): "A Comparision of Acoustic and Magnetic Resonance Imaging Techniques in the Measurement of Vocal Tract Shapes", proc. Fifth Irish Colloquium on DSP and Control, DCU Dublin, July 1994, pp. 27-34.

Djeradi, A., Guerin B., Badin, P., and Perrier, P. (1991). "Measurement of the acoustic transfer function of the vocal tract: a fast and accurate method". Journal of Phonetics **19**, 387-395.

Fant, G. (1970) " Acoustic theory of speech production". Mouton, the Hague. (1st edition 1960)

Fant, G. (1972). "Vocal tract wall effects, losses and resonance bandwidths", STL-QPSR 4/1985, pp. 1-13.

Fant, G. (1980). " The Relations between Area Functions and the Acoustic Signal", Phonetica 37, pp 55-86.

Fant, G., Nord, L., Branderud, P. (1976). " A Note on the Vocal Tract Wall Impedance". STL-QPSR 4/1976.

Fant, G., Lin, Q., and Gobl, C. (1985): "Notes on glottal flow interaction", STL-QPSR 2-3/1985, pp. 21-45.

Fant, G. (1989): "Speech Research in Perspective" proc. European Conference on Speech Communication and Technology, Vol. 1, pp 3-4. EUROSPEECH-89, Paris, Sept. 1989.

Fant, G. (1991) "What can basic research contribute to speech synthesis?", Journal of Phonetics **19**, 75-90.

Flanagan J.L. (1972). "Speech Analysis Synthesis and Perception" . Springer-Verlag. (1st edition 1965)

Flanagan, J.L., Ishizaka, K., Shipley, K.L. (1972) "Synthesis of Voiced Sounds from a Two-Mass-Model of the Vocal Cords", The Bell System Technical Journal, 50(6) pp. 1233-1268.

Flanagan, J.L., Ishizaka, K., Shipley, K.L. (1975): "Synthesis of speech from a dynamic model of the vocal cords and vocal tract," The Bell system technical journal 54(3), pp. 485-506.

Furui, S. (1989): Digital Speech Processing, Synthesis and Recognition, Marcel Dekker, INC., New York.

Fujimura, O. and Lindqvist-Gauffin, J. (1964): "On the sinewave response of the vocal tract", STL-QPSR 1/1964, pp. 5-10.

Fujumura, O. and Lindqvist, J. (1971), "Sweep-tone measurements of vocal tract characteristics", J. Acoust. Soc. Am. 49, pp. 511-558.

Greenwood, A.R, Goodyear C.C., and Martin P.A., "Measurements of vocal tract shapes using Magnetic Resonance Imaging.", IEE proceedings, Vol 139, No 6. 1992.

Guo, Q. and Milenkovic, P. (1993): "Optimal estimation of vocal tract area functions from speech signals constrained by x-ray microbeam data". Proc IEEE ICASSP. Volume 2, pp. 538-541.

Hogan, W.D. and Scaife R. (1994): "Measurements of the Acoustic Transfer Fucntion of the Vocal Tract using the Electromagnetic Shaker method of Djeradi et al". proc. Fifth Irish Colloquium on DSP and Control, DCU Dublin, July 1994, pp. 149-156.

Ishizaka, K., French, J., Flanagan, J. (1975): "Direct Determination of Vocal Tract Wall impedance", IEEE trans. ASSP-23(4), pp. 370-373.

Ishizaka, K., Matsudaira, M. and Kaneko, T. (1976): "Input acoustic-impedance measurement of the subglottal system" J. Acoust. Soc. Am. 60, pp. 190-197.

Jeevakumar, K., (1993): "Joint Estimation of Vocal Tract and Source Parameters of a Speech Production Model", unpublished Ph.D. thesis, Dublin City University, July 1993.

Keefe, D.H., Benade, A.H. (1981): "Impedance Measurement Source and Microphone Proximity Effects", J. Acoust. Soc. Am., 69(5), pp. 1489-1495.

Liljencrants, J. (1985): "Speech Synthesis with a reflection-type line analog", Doctoral Thesis, Royal Institute of Technology, Stockholm Sweden.

Lin, Qiguang. (1990): "Speech Production Theory and Articulatory Speech Synthesis", Doctoral Thesis, Royal Institute of Technology. Stockholm Sweden.

Lindblom, B. and Sundberg, J. (1971): "Acoustical consequences of lip, tongue, jaw and larynx movement", J. Acoust. Soc. Am. 50. pp.1166-1179.

Maeda, S. (1979): "An articulatory model of the tongue based on a statistical analysis.", J. Acoust. Soc. Am. 65, S1, p.22(A).

Maeda, S. (1982): "A Digital Simulation Method of the Vocal Tract System". Speech Communication 1 pp.199-229. North-Holland Publishing Company.

Maeda, S. (1989): "Compensatory Articulation in Speech: Analysis of x-ray data with an articulatory model", proceedings of EUROSPEECH-89, Paris 1989, Vol. 2, pp. 441-444.

Merhaut, J. (1968): "Method of Measuring the Acoustical Impedance", J. Acoust. Soc. Am. 45, p.331(A).

Mermelstein, P. (1967). "Determination of the vocal tract shape from measured formant frequencies", J. Acoust. Soc. Am. 41, pp. 1283-1294.

Mermelstein, P. (1973). "Articulatory model for the study of speech production", J. Acoust. Soc. Am. 53(4), pp. 1070-1082.

Rabiner, L.R. and Schafer, R.W. (1978): "Digtal Signal Processing of Speech Signals", Prentice-Hall, inc. Englewood Cliffs, New Jersey 07632.

Salava, T. (1988): "Acoustic Load and Transfer Functions in Rooms at Low Frequencies", Jnl. Audio Eng. Soc. 36(10), pp.763-775.

Scaife, R. and Kanagaratnam J. (1990): "Articulatory Speech Synthesis: Estimation of Data from Speech", First Irish Colloqiuium on DSP and Control, Queens University Belfast. pp. 19/1-19/7.

Schroeder M.R. (1984). "Number Theory in Science and Communication". Springer-Verlag, pp 274-276.

Schroeder. M.R. (1966): " Determination of the geometry of the Human Vocal Tract by Acoustic Measurements", J. Acoust. Soc. Am. 41 (4, pt. 2), pp. 1002-1010.

Schroter, J., Sondhi, M.M. (1989): "Dynamic programming search of artculatory codebooks", Proc. IEEE ICASSP, pp. 588-591.

Sondhi, M.M. (1979): "Estimation of vocal tract areas: The need for acoustical measurements" IEEE Transac. Vol. ASSP-27, No 3, pp. 268-273.

Sondhi, M.M. and Gopinath, B. (1971): "Determination of the vocal tract shape from the impulse response at the lips", J. Acoust. Soc. Am. 49, pp. 1867-1873

Sondhi, M.M., Resnick, J.R. (1983): "The Inverse Problem for the Vocal Tract: Numerical Methods, Acoustical Experiments, and Speech Synthesis", J. Acoust. Soc. Am., 73(3), pp 985-1002.

Sorokin, Victor N. (1992): "Determination of the vocal tract shape for vowels." Speech Communication 11 71-85. North-Holland.

Van den Berg, JW. (1958): "Myoelastic-Aerodynamic Theory of Voice Production". Journal of Speech and Hearing Research 1, pp. 227-244.

Appendix A

A 1.0 Derivation of the acoustic wave equations for a lossless tube.

The solutions to the wave equations (2.2a and 2.2b) for a small tube of cross-sectional area A have the form:

$$u(x,t) = u^{+}(t - x/c) - u^{-}(t + x/c) \qquad \text{(A1.1a)}$$

$$p(x,t) = \frac{\rho\, c}{A}\left[u^{+}(t - x/c) + u^{-}(t + x/c)\right] \qquad \text{(A1.1b)}$$

Where $u^{+}(t - x/c)$ represents a forward travelling wave and $u^{-}(t + x/c)$ represents a backward travelling wave. If the vocal tract is considered as a concatenation of such uniform lossless tubes with different cross sectional areas, the solutions of the separate sections combined using the equations of continuity and glottal and lip boundary conditions can be used to model the transfer function from the glottis to the lips [Rabiner & Schafer 1978]. The idea is illustrated in fig A1 below.

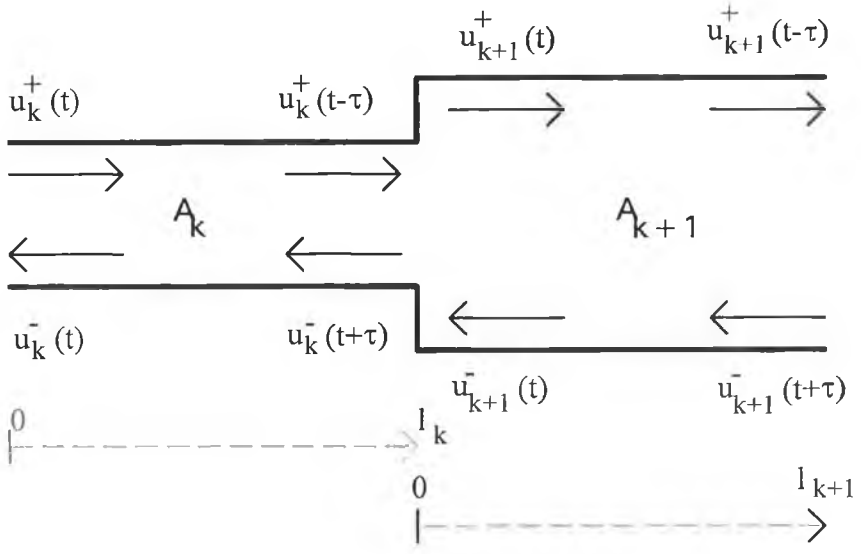


Figure A1: Travelling waves at the junction between two cylindrical tube sections

The pressure and volume velocity between the tube *k* and tube *k+1* are continuous across the boundary so we get:

$$p_k(l_k, t) = p_{k+1}(0, t) \qquad \text{(A1.2a)}$$

$$u_k(l_k, t) = u_{k+1}(0, t) \qquad \text{(A1.2b)}$$

From these equations it follows [Rabiner & Schafer '78]:

$$\frac{A_{k+1}}{A_k}\left[u_k^+(t-\tau_k)+u_k^-(t+\tau_k)\right]=u_{k+1}^+(t)+u_{k+1}^-(t) \tag{A1.3a}$$

$$u_k^+(t-\tau_k)-u_k^-(t+\tau_k)=u_{k+1}^+(t)-u_{k+1}^-(t) \tag{A1.3b}$$

This may be rewritten:

$$u_{k+1}^+(t)=(1+r_k)u_k^+(t-\tau_k)+r_k u_{k+1}^-(t) \tag{A1.4a}$$

$$u_k^-(t-\tau_k)=-r_k u_k^+(t-\tau_k)+(1-r_k)u_{k+1}^-(t) \tag{A1.4b}$$

where:

$$r_k=\frac{A_{k+1}-A_k}{A_{k+1}+A_k} \qquad -1\le r_k\le +1 \tag{A1.5}$$

is the reflection coefficient of the k^{th} junction, and represents the amount of the travelling wave reflected back at the junction. The time domain equations A1.4a and A1.4b can be converted into the digital z-domain as follows.

$$U_{k+1}^+(z)=(1+r_k)z^{-\frac{1}{2}}U_k^+(z)+r_k U_{k+1}^-(z) \tag{A1.6a}$$

$$U_k^-(z)=-r_k z^{-1}U_k^+(z)+(1-r_k)z^{-\frac{1}{2}}U_{k+1}^-(z) \tag{A1.6b}$$

where z^{-1} and $z^{-\frac{1}{2}}$ are delay elements in the digital domain equivalent to $-2\tau=-\frac{2l_k}{c}$ and $-\tau=-\frac{l_k}{c}$ respectively. Equation A1.6 can be written in matrix form as:

$$\mathbf{U}_k=\mathbf{Q}_k\mathbf{U}_{k+1} \tag{A1.7}$$

where $\mathbf{U}_k$ and $\mathbf{Q}_k$ are matrices given by:

$$\mathbf{U}_k=\begin{bmatrix}U_k^+(z)\\U_k^-(z)\end{bmatrix} \tag{A1.8}$$

$$\mathbf{Q}_k=\begin{bmatrix}\frac{z^{-\frac{1}{2}}}{1+r_k} & \frac{-r_k z^{-\frac{1}{2}}}{1+r_k}\\ \frac{-r_k z^{-\frac{1}{2}}}{1+r_k} & \frac{z^{-\frac{1}{2}}}{1+r_k}\end{bmatrix} \tag{A1.9}$$

The forward and backward travelling waves at each junction of the vocal tract system (not including the glottal and lip boundary conditions) can be described by the above equations. This allows us to compute the relationship between the input to the first tube $\mathbf{U}_1$ and the output at the beginning of the last tube $\mathbf{U}_N$ (N is the total number of tubes) by the matrix product:

$$\mathbf{U}_1 = \mathbf{Q}_1.\mathbf{Q}_2\cdots\mathbf{Q}_{N-1}.\mathbf{U}_N \qquad \text{(A1.10)}$$

A 1.1 Boundary conditions at the lips

At the lips (the end of the Nth tube) the flow of air is radiated as a pressure signal. The frequency domain relationship between the pressure and the flow signal is given as (for details of the equations below see [Rabiner& Schafer, 1978]):

$$P_N(l_N,\Omega) = Z_L U_N(l_N,\Omega) \qquad \text{(A1.11)}$$

where Z_L is defined as the radiation impedance or the radiation load at the lips. It has been shown that the matrix relating the flow in the Nth tube to the flow at the lips is given by:

$$\mathbf{U}_N = \begin{bmatrix} \dfrac{z^{\frac{1}{2}}}{1+r_L} \\ \dfrac{-r_L z^{-\frac{1}{2}}}{1+r_L} \end{bmatrix}.\mathbf{U}_L \qquad \text{(A1.12)}$$

where:

$$r_L = \left[\frac{\frac{\rho c}{A_N} - Z_L}{\frac{\rho c}{A_N} + Z_L}\right] \qquad \text{(A1.13)}$$

A 1.2: The boundary condition at the glottis.

The model must take into account the different sound excitation sources. The three main types of excitation are voiced, fricative, and voiced frication. During the production of vowel sounds the vocal tract enters a condition of sustained oscillation. The coupling between the vocal tract and the glottis is weak, and one can neglect their interaction and

treat the excitation system as linearly separable from the transmission system [Flanagan, Ishizaka and Shipley, 1972]. Note: Source-tract interaction is considered in more complex modelling e.g. [Fant, Lin and Gobl, 1985; Lin, 1990]. In natural speech the mechanism of production of fricative sounds involves the turbulent flow of air. This can occur at a constriction whenever the volume flow exceeds a certain critical value. Such excitation can be modelled by the insertion of a randomly time-varying source at the point of constriction. For voiced fricatives the strength of a frication source can be made dependent on the volume velocity in the tube and in this way frication is automatically inserted when needed e.g. [Flanagan, Ishizaka and Shipley, 1975]

Assuming that the excitation source at the glottis is linearly independent from the vocal tract, the flow in the first tube is given by [Rabiner & Schafer, 1978]:

$$U_1(0,\Omega) = U_G(\Omega) - P_1(0,\Omega)/Z_G \qquad \text{(A1.14)}$$

where U_1 and P_1 are the pressure and volume velocity at the input of the first tube and Z_G is the acoustic impedance of the glottis. Rewriting this equation in terms of backward and forward travelling waves (eqns A1.1) we get:

$$u_1^+(t) = \frac{(1+r_G)}{2} u_G(t) + r_G u_1^-(t) \qquad \text{(A.1.15)}$$

where the reflection coefficient at the glottis is given by:

$$r_G = \left[\frac{Z_G - \frac{\rho c}{A_1}}{Z_G + \frac{\rho c}{A_1}} \right] \qquad \text{(A1.16)}$$

In the digital domain with a little rearranging this relation is written in matrix form as:

$$\mathbf{U}_G(z) = \left[\frac{2}{1+r_G}, \quad \frac{2r_G}{1+r_G} \right] . \mathbf{U}_1 \qquad \text{(A1.17)}$$

A 1.3 Complete equation of the vocal tract model

The transfer function of the complete vocal tract model may now be obtained by combining the lip and glottis boundary conditions to the tube model (equations A1.10, A1.12 and A1.17). This gives us the following equation:

$$U_G(z) = \left[\frac{2}{1+r_G}, \quad \frac{2r_G}{1+r_G}\right] \cdot \mathbf{Q}_1 \cdots \mathbf{Q}_{N-1} \begin{bmatrix} \frac{z^{\frac{1}{2}}}{1+r_L} \\ \frac{r_L z^{-\frac{1}{2}}}{1+r_L} \end{bmatrix} U_L(z) \qquad \text{(A1.18)}$$

The transfer function of the model $V(z)$ relates the output flow at the lips to the input flow at the glottis and is given by:

$$V(z) = \frac{U_L(z)}{U_G(z)} = \frac{0.5(1+r_G)(1+r_L)\prod_{k=1}^{N-1}(1+r_k)z^{\frac{-N}{2}}}{D(z)} \qquad \text{(A1.19)}$$

where the denominator $D(z)$ is given by:

$$D(z) = \left[1, \quad -r_G\right] \cdot \mathbf{Q}_1 \cdots \mathbf{Q}_{N-1} \begin{bmatrix} 1 \\ -r_L z^{-1} \end{bmatrix} \qquad \text{(A1.20)}$$

See chapter 4 for measurements of transfer functions for vowels and comparisons to a model. The transfer function given in equation A1.19 is that of the basic lossless digital model. A lossy digital speech production model is described in [Liljencrants, 1985]. He has added series and shunt components to the model to take into account viscous friction, heat conduction and wall vibration losses.

Appendix B

B 1.0 The analog interface

The LSI TMS320C30 system boards have a "complete analog subsystem", consisting of two channels each with its own sample/hold amplifier, A/D, D/A and analog filters on input and output. The A/Ds and D/As are Burr Brown devices which offer 16-bit precision (theoretically; see section 3.7) with up to 200kHz sampling rates. The analog subsystem is accessed through three 32-bit I/O bus mapped registers; only the top 16-bits of each register are used.

Function	Address
R/W Channel A A/D and D/A	804000H
R/W Channel B A/D and D/A	804001H

Table B.1: Analog I/O mapping

For example to output data to the D/A on channel A, you move data to the I/O-mapped register located at 804000 hex, and to input data on the same channel, you get data from the same register. The fact that both input and output are performed using the same register means that for simultaneous input/output (reading and writing samples of data in the same sampling period), the register must always be read before it is written to, otherwise you end up reading the sample you have just written out.

The A/D and D/A conversions on both I/O channels are initiated by an on-chip counter/timer (Timer 1) of the 'C30. Timer 1 comprises a 32-bit up counter and a 32-bit period register. The value in the counter is continuously incremented at a rate of 8.33MHz, when this value equals the value in the period register, a pulse signal resets the counter register and is outputted to the A/D and D/A converters to start a conversion. The data from the A/D and D/A converters is dealt with using interrupt service routines. After an A/D conversion, the converter outputs an end of convert signal, which is linked to the 'C30 as the INT1 interrupt request. When this interrupt is enabled, the software interrupt

service routine can immediately read one or both A/Ds, and then if desired write to one or both D/As. The converters deal with data in 16-bit 2's complement format, which is also the format of 'C30 integers. Unfortunately there are limitations on the bit resolution of the on-board converters when used in a real system; our solution to this problem is discussed in the next section.

B 1.1 The serial-ports of the TMS320C30

The TMS320C30 has two totally independent bidirectional serial-ports. Both serial-ports are identical, and can be configured to transfer 8, 16, 24 or 32 bits of data per word simultaneously in both directions. The clock for each serial-port can originate either internally, via the serial-port timer and period registers, or externally, via a supplied clock.

Each of the serial-ports is provided with eight memory mapped registers:

- Global-control register
- Two port control registers for the six serial I/O pins
- Three receive/transmit timer registers
- Data-transmit register (DXR)
- Data-receive register (DRR)

The global-control register is a 32-bit register that contains the global control bits for the serial-port. These control fuctions such as specifiying the type of clock to be used, and resetting of the serial-port. There are two port control registers, one to control the serial-port transmit pins (FSX, DX, and CLKX) and the other to control the serial-port receive pins (FSR, DR, and CLKR). The three receive/transmit (r/t) timer registers are the r/t timer control register, the r/t timer period register, and the r/t timer counter register. These registers are used to generate a software configurable data clock.

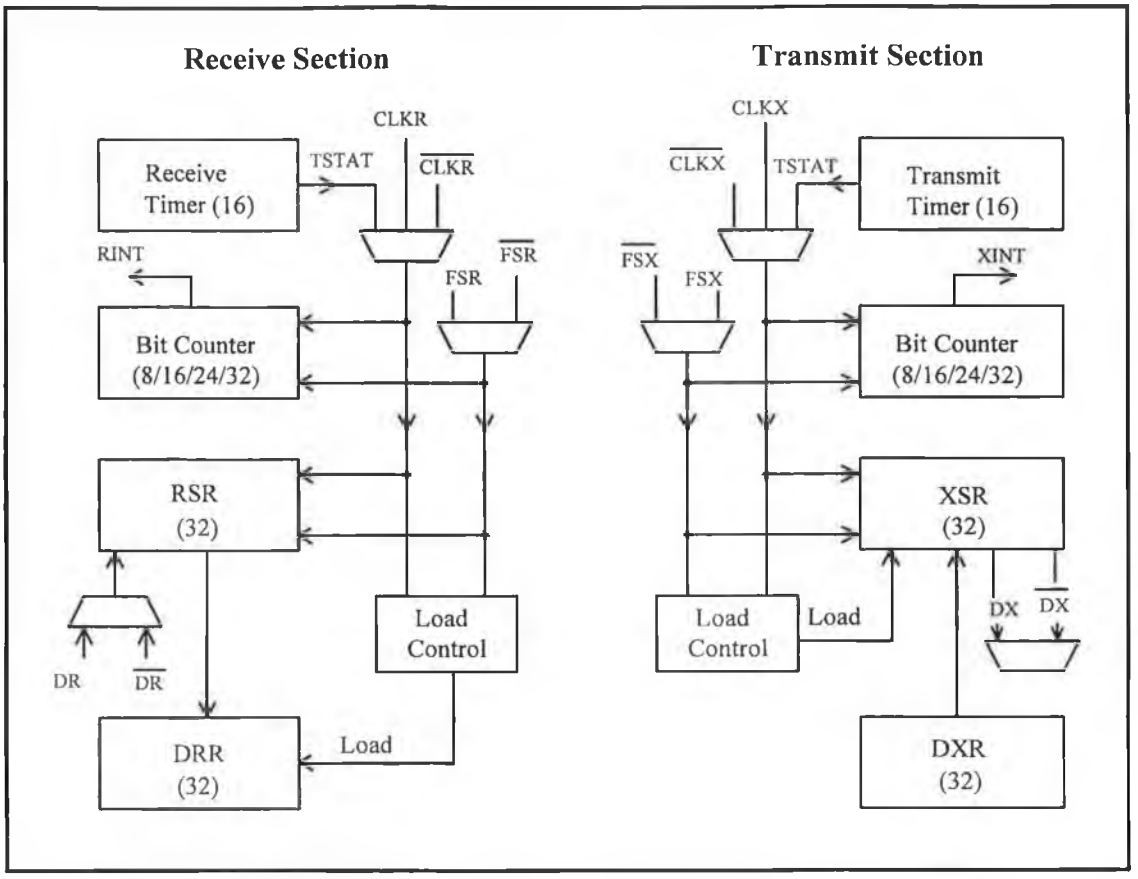


Figure B 1.1: Block diagram of the serial-port

A sample to be transmitted is loaded into the data-transmit register (DXR), and from there it is loaded into the transmit-shift register (XSR), see figure B.1.1. A sample being received is loaded into the receive-shift register (RSR), and from there is loaded into the data-receive register (DRR). The XSR and the RSR are not memory mapped and so are not directly accessible from software.

B 1.2 Serial-port interrupt sources:

Each serial-port has four interrupt sources. These are:

1. The transmit timer interrupt: The rising edge of XTSTAT causes a single cycle interrupt to occur. When XTINT is 0, this interrupt is disabled.
2. The receive timer interrupt: The rising edge of RTSTAT causes a single cycle interrupt pulse to occur. When RTINT is 0, this interrupt pulse is disabled.

3. The transmitter interrupt: Occurs immediately following a DXR-to-XSR transfer. The transmitter interrupt is a single cycle pulse. When the serial-port global-control register bit XINT is 0, this interrupt is disabled.
4. The receiver interrupt: Occurs immediately following a RSR to DRR transfer. The receiver interrupt is a single cycle pulse. When the serial-port global-control register bit RINT is 0, this interrupt pulse is disabled.

The transmit timer interrupt pulse is ORed with the transmitter interrupt pulse to create the CPU transmit interrupt flag XINT. The receive timer interrupt is ORed with the receiver interrupt pulse to create the CPU receive interrupt flag RINT. For the present project 3 and 4 above were used to generate interrupts. This in effect means that the interrupts are synchronised to the sampling rate. These interrupts are used to initiate software interrupt service routines (ISRs) to control the transmission and reception of data. The interrupt service routines also specify the other operations of the TMS320C30 between the transfer of successive samples. The ISRs are explained in greater detail in B 2.1.

B 1.3 Serial-port configuration:

All of the serial-port operating configurations can be classified into two broad categories: fixed data-rate timing and variable data-rate timing. Fixed data rate timing can occur in two modes; burst mode and continuous mode. In burst mode operation transfers of single words are separated by periods of inactivity on the serial-port. In continuous mode there are no gaps between successive word transfers; the first bit of a word is transferred on the next bit-clock pulse following the last bit of the previous word, this continues until the process is terminated. Fixed burst mode timing was chosen for the present project. With a typical audio sampling rate of 16kHz and a bit clock of about 8MHz, the serial-ports only need to be active for less than 10% of the time between the transfer of two successive samples. In fixed burst mode the rising edge of the FSX/FSR clock (the sampling rate clock) initiates transfers and each transfer involves a single 32-bit word. In a transmit operation, the FSX pulse initiates a transfer from the DXR to the XSR (see figure B 1.2).

There is a setup requirement that DXR must be loaded no later than 3 CLKX cycles before FSX occurs. Once the XSR is loaded from the DXR an XINT is generated.

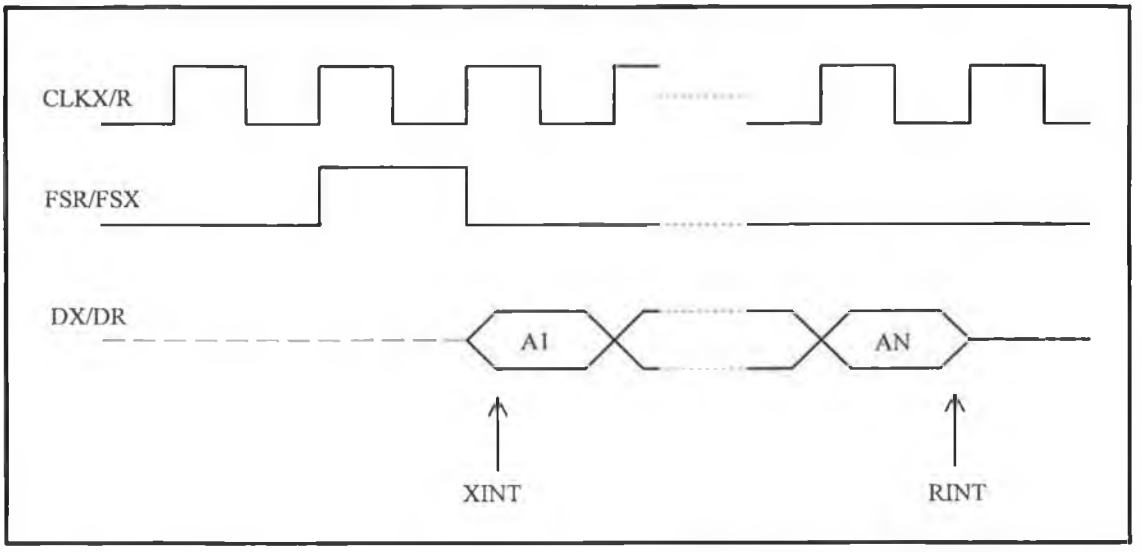


Figure B 1.2: Fixed burst mode timing diagram

In a receive operation the rising edge of FSR initiates a transfer from the RSR to the DRR. Once a transfer is initiated, FSR is ignored until the last bit has been received. FSR must be low during the last bit or another transfer will be initiated. When a full word has been received and transferred to the DRR, an RINT is generated.

B 1.4 Operation of the interface circuit.

The serial-port interface circuit is based around two high performance ICs from Burr-Brown, the DSP102 analog-to-digital converter, and the DSP202 digital-to-analog converter. Both of these ICs have complete logic interface circuitry for ease of use with standard digital signal processing ICs and transmit data words in a serial stream. The DSP102 and the DSP202 are 18-bit devices internally. When the interfaced DSP IC is programmed to accept 16-bit word lengths, the processor ignores the last two bits received from the DSP102 and transmits only 16-bits to the DSP202. The converters are two channel devices that can operate in a number of different modes. For data transfer these modes are:

- Two separate 18-bit channels
- Two separate 16-bit channels
- Cascade mode: The 16 bits of each channel are packed into a 32 bit word before transmission along one of the channels.

The DSP102 A/D is interfaced to the TMS320C30 serial-port receive side, the DSP202 D/A is interfaced to the transmit side. There are a number of clock, or timing signals to be considered in the design. These are:

- The data transfer or bit-clock
- The conversion clock
- The sampling rate clock

The bit clock is the rate at which each bit of serial data is clocked into and out of the serial-port. It is generated by the CLK0 pin on the serial-port, and fed into the XCLK pins on the converters.

The conversion clock determines the internal speed of the converter ICs, and therefore the amount of delay between the reception of a convert command to the time the converted data is ready to be transferred. The conversion clock may be derived from the bit clock by feeding the bit clock into the OSC1 pin of the DSP102. The signal is divided by three internally, outputted to the CLKOUT pin of the DSP102 and fed back into the CLKIN pin. This method avoids the use of external crystal oscillator circuits, and also synchronises the bit clock and the conversion clock, minimising glitches.

The sampling rate clock controls the sampling rate of the system. The Burr Brown ICs can run at a maximum sampling rate of 150kHz if all the other clocks are at their maximum speed. It is desirable to have both the A/D and D/A synchronised to the same sampling rate as this will minimise distortion and also allow greater control over simultaneous record and playback. The sampling rate clock is connected to the CONV pin on the DSP102 and the DSP202.

B 1.5 The sampling rate clock

At this point a serious design problem was encountered. It is desirable to have a software configurable sampling rate clock, to allow a flexible sampling rate which can be easily changed by just modifying a line of code or a setup parameter. This is generally done by using one of the two on-chip timers on the TMS320C30. The programmed timer clock is outputted to the TCLK0 or the TLCK1 pin on the TMS320C30 and from there fed to the A/D and D/A converters. Unfortunately on the LSIC30 board neither of the timer pins on the TMS320C30 are fed to the serial endplate connector, and thus are unavailable. Soldering wires directly on to the LSIC30 board was not considered a viable solution because it was necessary to have a portable system that could plug into any LSIC30 board.

It seemed initially that the only alternative was a system of external oscillators covering the most essential sampling rates using hardware jumpers or a switch to select between them. Fortunately there was a solution to this problem because of the special cascade mode mentioned earlier. There are two serial-ports on the TMS320C30; serial-port 0 and serial-port 1. In separate channel mode both of the serial-ports are used to transfer two channels of data. However in cascade mode only serial-port 0 is used to transfer data; channel A data is stored in the top 16 bits and channel B data is stored in the bottom 16 bits of a 32-bit data word. In this mode serial-port 1 is free. This allows the serial-port timer of serial-port 1 to be used to generate the sampling rate clock, instead of doing its usual job of generating a bit rate clock. The serial-port timer can be programmed to work at a wide range of frequencies including the complete sampling rate range of the converters. The serial-port timer works by using a timer register and a period register. The timer register increments at about a speed of about 8.3MHz, and when it reaches the value in the period register it is reset. The value placed in the period register determines the sampling rate.

Of course, as the values loaded into these registers are discrete, so are the sampling rates available. For instance, a counter load value of 189 gives a sampling rate of 44.09kHz, and a counter load value of 188 gives a sampling rate of 44.3kHz. The discrete nature of

the sampling rate clock did not pose a problem for the purposes of the measurements presented in this thesis. This cascade mode solution allowed the serial interface to connect to the LSIC30 using only the serial endplate connector (DB15) shown in fig. B1.3.

The circuit setup is shown in fig B1.3 (this is the actual ORCAD diagram used to help generate the PCB layout). The DSP102 A/D is interfaced to the serial-port receive side, the DSP202 D/A is interfaced to the transmit side. The A/D and D/A are hardwired to run in cascade mode. When the TMS320C30 initiates a convert command to the A/D via the CLKX1 pin of serial-port 1, both analog inputs are converted into two 16-bit words which are concatenated to form one 32-bit word. The A/D signals the TMS320C30 via the A/Ds SYNC signal (connected to the TMS320C30 FSR0 pin), that serial data is to be transmitted. The 32-bit word is then serially transmitted MSB first, out the SOUTA serial pin of the DSP102 to the DR0 pin of the TMS320C30 serial-port. The TMS320C30 is programmed to drive the bit clock from the CLK0 pin of serial-port 0.

Similarly upon receiving a convert command, the D/A converts the last word received from the TMS320C30 and signals the TMS320C30 via the SYNC signal (connected to the TMS320C30 FSX0 pin), to begin transmitting a 32-bit word representing the two channels of data to be converted. The data transmitted from the TMS320C30 DX0 pin is input to both the SINA and SINB inputs of the D/A. The two channels are seperated, and on the next convert command converted and outputted to the DSP202 analog output pins VOUTA and VOUTB.

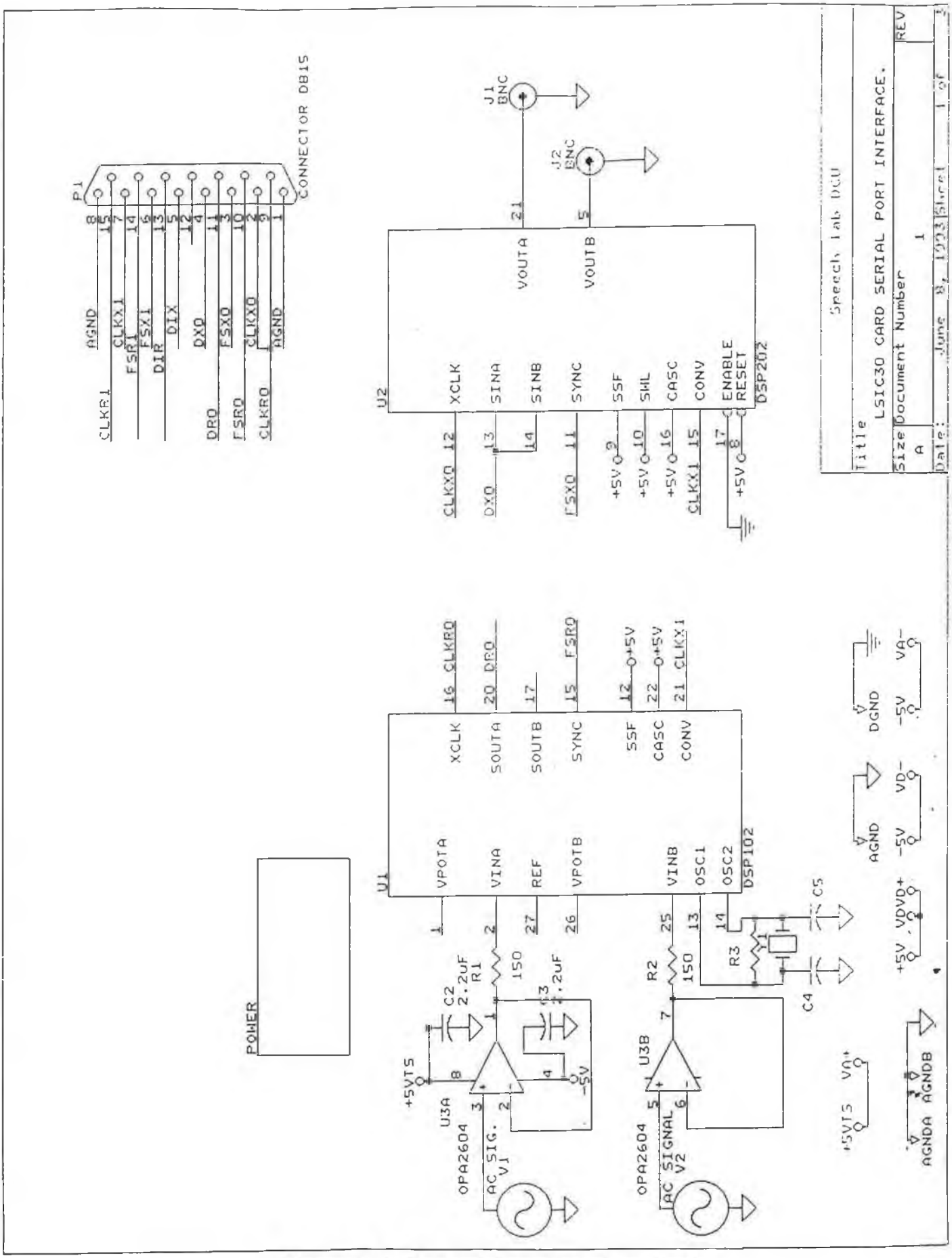


Figure B1.3: The serial port interface circuit

B 1.6 Circuit design and construction

The next stage of the design was to consider the analog inputs and outputs, and the power supplies. The DSP102 A/D requires its analog inputs to be driven by sources with low impedances over the input bandwidth used in the application. Standard 74 series op-amps are used to buffer the inputs and satisfy this criterion. The 150Ω and 220pF shown on the analog inputs in figure B1.3 help reduce the dynamic load on the input signal conditioning amplifier in front of the A/D. Buffers are not required on the analog outputs of the DSP202 D/A as they are provided internally.

The system design requires three power supplies. A +5V digital power supply, a +5V analog power supply, and a -5V analog power supply. All these supplies must be decoupled to the appropriate ground using tantalum capacitors in parallel with ceramic capacitors. The analog and digital supplies do not necessarily have to be separate but it is recommended that they are. When they are separate there should never be a difference of more than 0.5V between them, except for a few hundred milliseconds as may occur at power-up. Low power voltage regulators (LP2950CZ) were added to the circuit to keep the supply voltages constant.

B 1.6.1 Testing the circuit

The circuit was designed and a circuit diagram was drawn using the ORCAD package (fig. B 1.3). The circuit was built on a breadboard and tested. When the testing was complete and all the necessary modifications had been made a PCB layout was designed.

The breadboard circuit was tested using a C-language program that continuously read a sample from the DSP102 into the serial-port receive side and echoed the sample out to the DSP202 via the serial-port transmit side. The testing results showed that approximately 11-bits accuracy could be obtained from the breadboard design. This was 1-bit better than the onboard converters, and it was expected that a PCB design would improve this further.

B1.6.2 PCB design and manufacture

The PCB layout was designed using the Boardmaker package from Tsien ltd. This package provides a full PCB schematic editor that allows the design of boards with up to 8 layers. The package has a library of standard IC sizes for ease of use, and also allows a wide variety of track thicknesses and pad sizes. When the layout design is complete the package carries out a series of design rule checks on such factors as track-to-track distances and track-to-pad distances; any violations may be modified. The finished design may be output in Gerber format and NC drill format which can be used directly by a printed circuit board manufacturer.

The Burr-Brown converters are very sensitive devices. Conversion accuracy can easily be corrupted by such factors as current flow in the system grounds. It is recommended in the Burr-Brown data sheets that wherever possible separate analog and digital ground planes should be used, as the currents switched in a typical DSP system can be quite large. For this reason it was decided to design a four layer PCB with separate power and ground layers, and two connection layers. The ground layer was further split into an analog ground grid and a digital ground grid with one central connection between them. It is very important to ensure there are no errors in the PCB layout with this type of board, as it is very difficult to modify the two internal layers once the board is manufactured. When the design was complete the Gerber output files were sent to an external manufacturer, and 10 copies of the board were made. The components were soldered on to the board and the testing procedure was carried out as before. The testing showed that the bit resolution was now 13-bits, which was an improvement of 2-bits on the breadboard construction, and 3-bits on the onboard converters.

When the circuit was fully tested and found to be working, all that remained was to place the whole system in a housing box and add on all the external connectors. The housing box chosen was zinc alloy with dimensions $140 \times 100 \times 75mm$. It was decided to allow a choice between battery power and external (benchtop) power supplies to keep the system as portable as possible. To this end a switching system was designed for a 4 pole three

way switch, connected to the outside of the housing case. The three modes provided for were.

- Off/External power enable.
- Internal (Battery) power enable.
- Battery charge mode.

The batteries chosen were 9V PP3 type rechargables, and five were required. The third mode was added to allow the batteries to be recharged from a standard benchtop power supply without removing them from the housing box. Four BNC posts were added for the analog inputs and outputs, and a DB15 type connector was added to interface to the serial-port endplate connector of the LSIC30 board.

B 2.1 Methods of programming the LSIC30 board.

B 2.1.1 Assembly based

DSP applications can be written in TMS320C30 assembly code, assembled and linked using the Texas Instruments assembler and linker tools, and loaded down to the board using a debug monitor (SDS30) or by writing a C program that uses library functions to load and run the assembled application code. The debug monitor displays the register values on the system board, and allows single stepping through the code to check its operation.

B 2.1.2 Plain C based

The only difference between this method and the first is that the applications are written in C initially, and compiled using the Texas Instruments C compiler and linker. This automatically converts the C programs into assembly programs, and then assembles them to give executable dsp files. However, it was found that writing all of the application programs in C is not possible. Assembly language routines have to be included in the C program to deal with certain types of register manipulation and interrupt service routines. There is also the problem that the assembly code generated by the compiler is generally quite inefficient.

B 2.1.3 SPOX based

Spox is a higher level DSP operating system that provides a set of pre-written optimised functions that can simplify the creation of DSP algorithms. It provides functions that take care of input/output streaming, fourier transforms, filters etc. The applications are written in C (ANSI standard) and are compiled, assembled and linked using the cl30 tool. They can be loaded down to the board and run using the EXECLSI tool, or with a C program using the LSI libraries as before. All the SPOX functions rely on arrays. A simultaneous input to the A/D converter, and output to the D/A converter cannot be performed because the incoming data has to be read into an array, and when the array is full, the data can be written out to the D/A. The instream function (GET) cannot be interrupted until it has filled a buffer or there are no more samples to be read in, so an outstream (PUT) funtion cannot run at the same time.

B 2.2 TMS320C30 assembly language programming

All the assembly language programs used for this project have the same general format.

1) Processer initialisation
2) Signal input/output
3) Data Processing

These separate functions are described briefly in the next three sections.

B 2.2.1 Processor initialisation:

Prior to the execution of a DSP algorithm it is necessary to initialise and configure the processor. Initialisation occurs when the processor is reset. To configure the processor after a reset the following internal functions should be initialised.

- Memory mapped registers
- Interrupt structure
- Memory configuration
- On chip timers
- Data memory page pointer

- Serial ports (if required)

B 2.2.2 Signal input/output

For our purposes simultaneous data I/O was required. Every sampling period a sample must be written from board memory to the D/A converter, and in the same sampling period a sample must be read from the A/D converter and stored in the board memory of the LSIC30. This is done using an interrupt service routine (ISR) synchronised to the sampling rate via the on-chip timers, or in the case of the serial port system via the serial port timer. When an enabled interrupt occurs, the CPU halts all activites, disables further interrupts, and sets the program counter to the address of the ISR found in the interrupt vector table. When it has processed the ISR commands it returns the program counter to its previous value and re-enables interrupts. For the dual transfer function/lip impedance measurements described in chapter 5, two interrupt service routines were required; one initialised from the serial-port system, and the other from the on-board converter system.

B 2.2.3 Data processing

Once the signal has been acquired further processing can take place. For the transfer function measurements described in chapter 4, the processing consisted of a circular cross-correlation routine that required approximately 12 million multiply and add operations. The TMS320C30 has a number of features that can decrease program execution time if taken advantage of properly. These include:

- Parallel instructions to increase the number of instructions executed in a single cycle.
- Repeat block instructions: Normal branch instructions have an overhead of 4 cycles. Repeat block instructions implement zero overhead looping and preserve the pipeline.
- Use of the register file instead of accessing slow external memory.
- Use of the cache.

Use of the features above provided a significant decrease in the execution times for the measurement programs.

B 2.3 Testing the I/O system:

When the I/O system was built it was tested using a simple echo program. A flowchart for the on-board section of this program is given below in figure B2.2 In this diagram XINT represents a transmit interrupt. A file of pseudo-random samples was read out to transmit port of the serial port. The serial port transmit side was connected externally to the serial port receive side. The samples were read in to the serial port receive side and stored in memory. Comparison of the two files using cross-correlation techniques (described in Chapter 4) showed exact correspondence with a noise resolution of 13 bits.

B2.4 The theory of generation of pseudo-random sequences

A shift register is a collection of storage elements connected so that the state of each element is shifted to the next element in response to a shifting clock or timing signal. A linear binary network is one constructed from the following basic components.

- Unit delays
- Modulo 2 adders
- Modulo 2 scalar multipliers

A network constructed of linear elements has the property that its response to a linear combination of inputs will preserve the principle of superposition. A shift register with a linear feedback network is called a linear feedback shift register (LFSR). A linear feedback shift register is a finite state machine. Each state is uniquely determined from the previous state by the feedback connections. If a state ever repeats, all the following states will repeat, and the sequence of states is periodic. Given an n-stage shift register, there are at most 2^n possible states. With an initial state of all zeros, all the following states are all zero, as the outputs of the feedback elements are zero when all their inputs are zero. Hence, an LFSR can hold at most $2^n - 1$ possible states before repeating. If the

sequence generated by an n-stage LFSR has period $2^n - 1$ it is called a maximal length sequence, or m-sequence. A *binary* sequence can be generated from the LFSR by taking the contents of any one of the storage elements as each of the states are shifted through. The period of the binary sequence is the same as the period of the states.

A three stage LFSR that generates an m-sequence is shown in figure B2.1 below. The + sign represents modulo 2 addition (XOR). In this case there are $2^3 - 1 = 7$ states. The m-sequence is obtained by taking any of the three columns shown in figure B2.1. If we take the third column then the sequence is $a_m = \{0010111\}$. The sequence a_m is cyclic in that it will repeat every 7 cycles of the shift register. These types of sequences can be represented mathematically by their generating function. The generating function for this type of sequence can be represented by its characteristic polynomial. For the example shown in figure B2.1 the characteristic polynomial is given by:

$$f(x) = 1 + x + x^2 \qquad \text{(B 2.1)}$$

We see that the 'powers' in this equation represent the feedback taps shown in figure B2.1. The characteristic polynomial of a maximal length sequence is called a primitive polynomial. Given the primitive polynomial for a sequence of degree n, we can construct the shift register necessary to generate that sequence. Tables of primitive polynomials for degrees of up to 300 are available. To generate a maximal-length binary sequence of length 2047, we need a primitive polynomial of degree 11; this turns out to be:

$$f(x) = 1 + x^2 + x^{11} \qquad \text{(B 2.2)}$$

The primitive polynomial used to generate a sequence of length 4095 is:

$$f(x) = 1 + x^3 + x^4 + x^7 + x^{12} \qquad \text{(B 2.3)}$$

A twelve stage shift register is used in the above case.

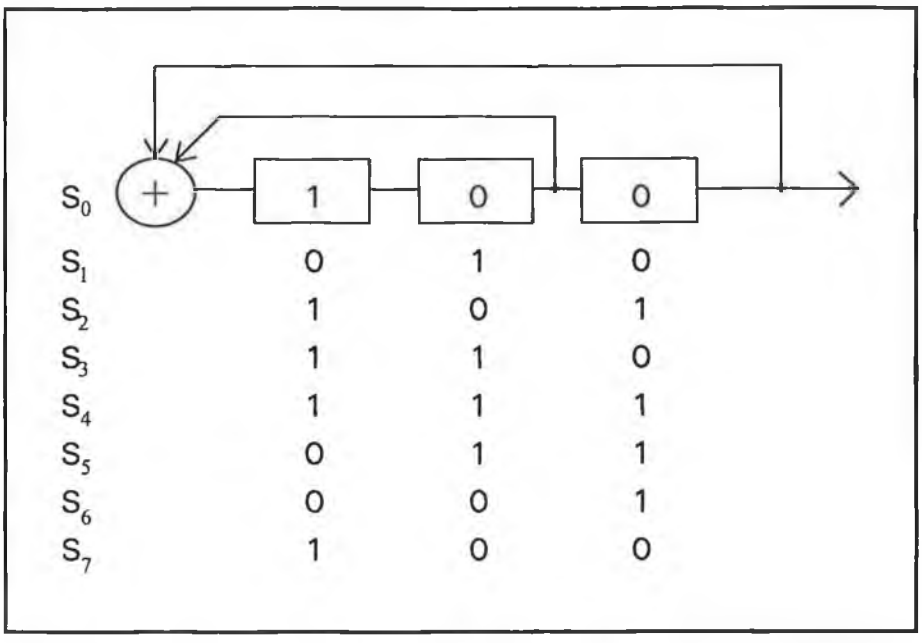


Figure B 2.1: Three stage LFSR to generate an m-sequence

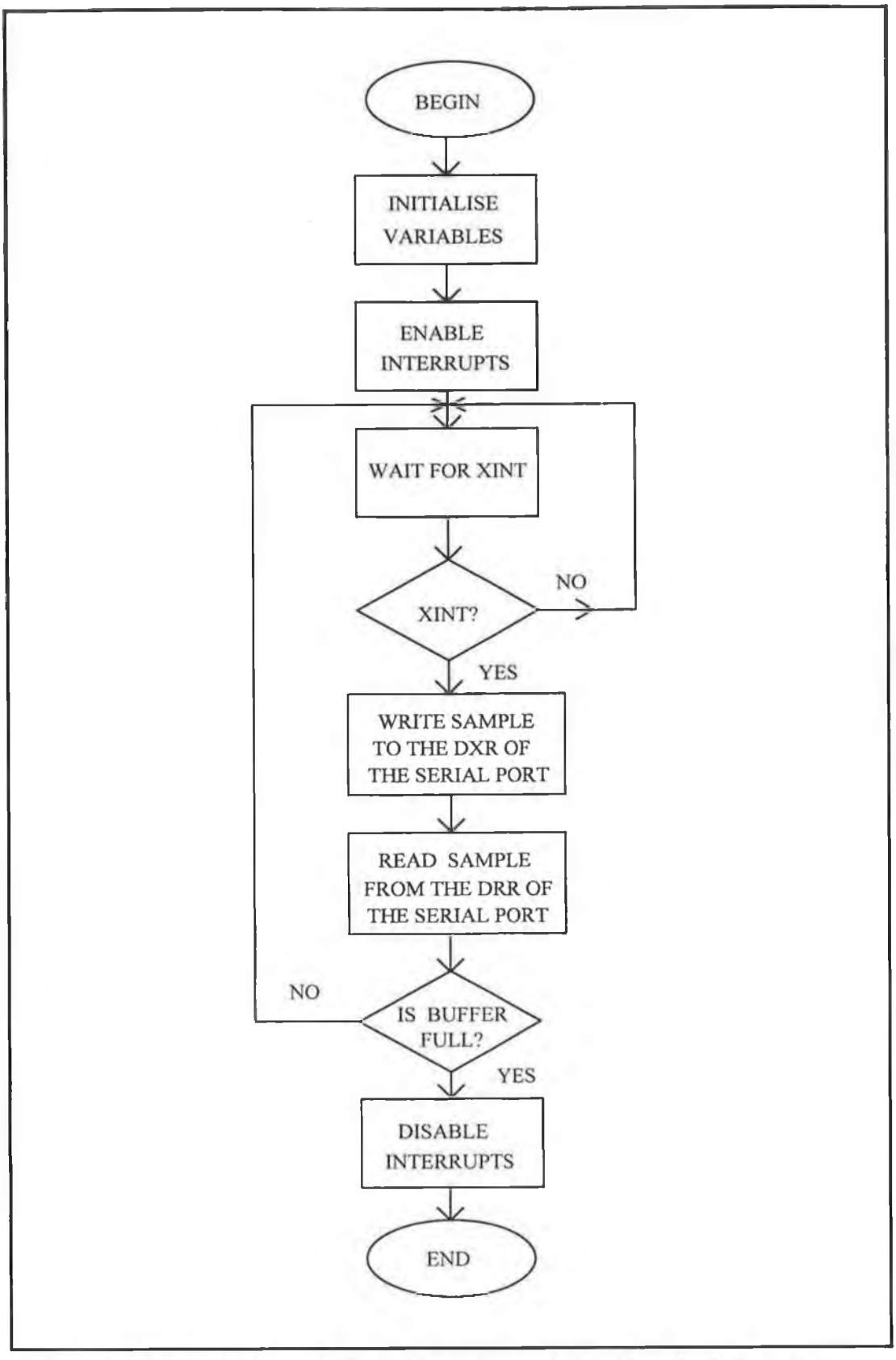


Figure B2.2: Flowchart of the echo program for testing the I/O system

Appendix C

C 1.0 Correction for the lip adaptor tube

For measurements on human subjects, the impedance measurement head was connected to a variety of lip-adaptor tubes. The lip-adaptor tubes are copper pipes with one end of the same cross-sectional area as the measurement head, and the other end at a narrower cross-sectional area to couple to the subject's lips, and allow the subject to form a natural and comfortable articulation. For speech measurements it was necessary to correct the measured data to eliminate the impedance of the adaptor tube, and arrive at the acoustic impedance seen at the lips. The correction was performed by use of the lumped equivalent circuit of the adaptor tubes. Figure C1 below shows the equivalent circuit for a general adaptor tube.

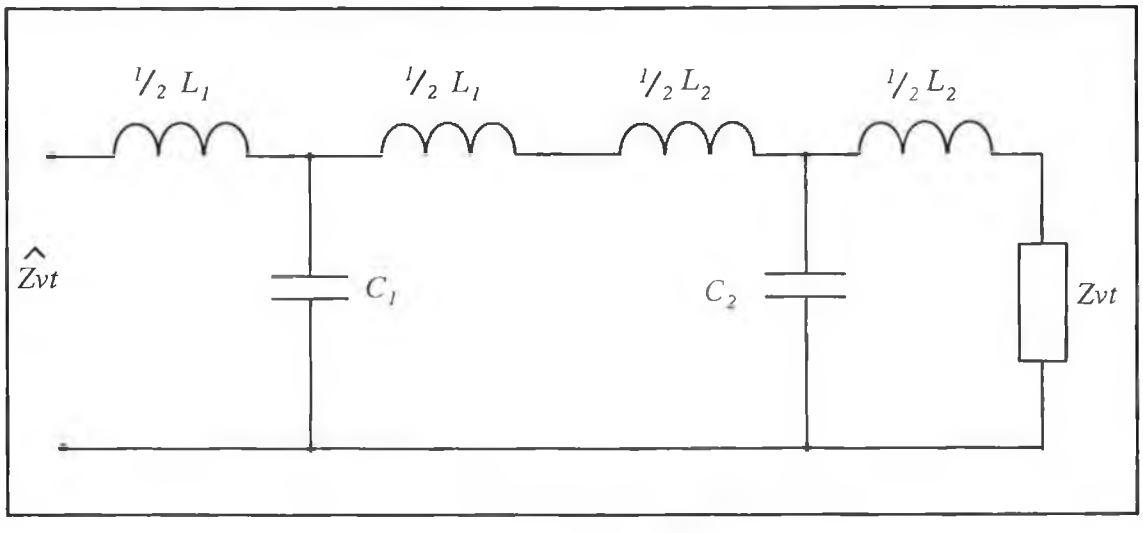


Figure C1: Equivalent circuit for a general lip adaptor tube

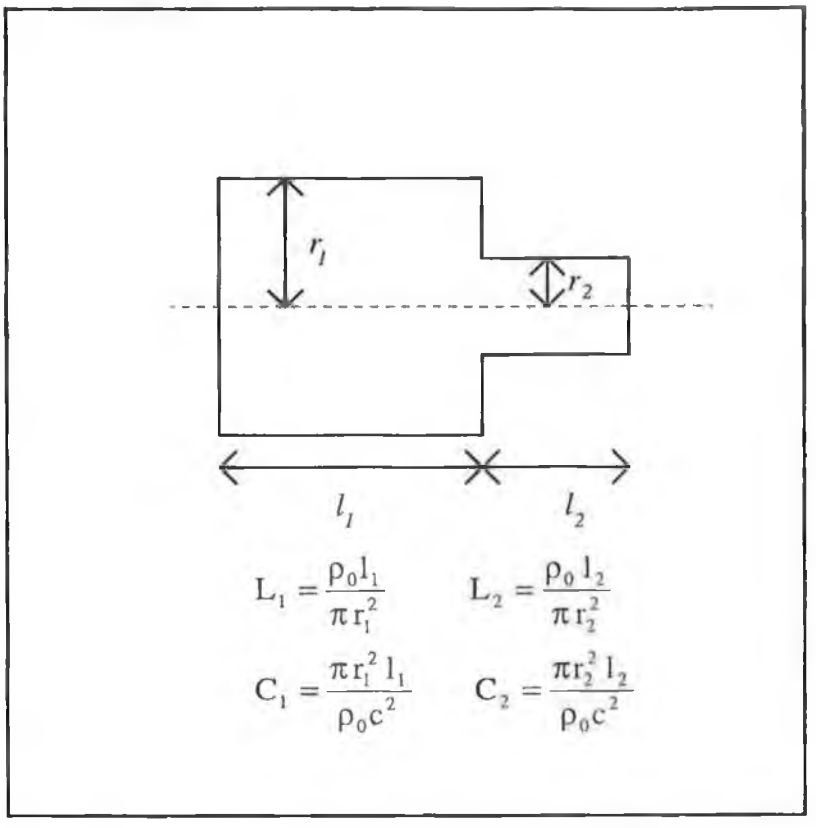


Fig C2: Relationship between a lip adaptor tube and its equivalent circuit

The desired impedance at the lips, Z_{vt}, is obtained from $\hat{Z}_{vt}$, the measured impedance, by progressively subtracting[1] the values of the equivalent circuit components $1/2\,L_1$, C_1, $1/2(L_1 + L_2)$, C_2 and $1/2\,L_2$ at each frequency as follows:

$$Z_{vt,1} = \hat{Z}_{vt} - \frac{1}{2} j\omega L_1$$

$$Y_{vt,1} = \left(Z_{vt,1}\right)^{-1}$$

$$Y_{vt,2} = Y_{vt,1} - j\omega C_1$$

$$Z_{vt,2} = \left(Y_{vt,2}\right)^{-1}$$

$$Z_{vt,3} = Z_{vt,2} - \frac{1}{2} j\omega\left(L_1 + L_2\right)$$

$$Y_{vt,3} = \left(Z_{vt,3}\right)^{-1}$$

[1]The impedance is converted to an admittance before subtracting the capacitance values

$$Y_{vt,4} = Y_{vt,3} - j\omega C_2$$

$$Z_{vt,4} = \left(Y_{vt,4}\right)^{-1}$$

$$Z_{vt} = Z_{vt,4} - \frac{1}{2} j\omega L_2$$

This correction is only valid for wavelengths much greater than the typical lengths of the sections. For the measurements presented below we only require that the correction be satisfactory up to about 1kHz, where the wavelengths are more than ten times greater than the 25mm length of the adaptor elements. For higher frequency corrections, it would be necessary to split the adaptor model into shorter cylindrical elements.

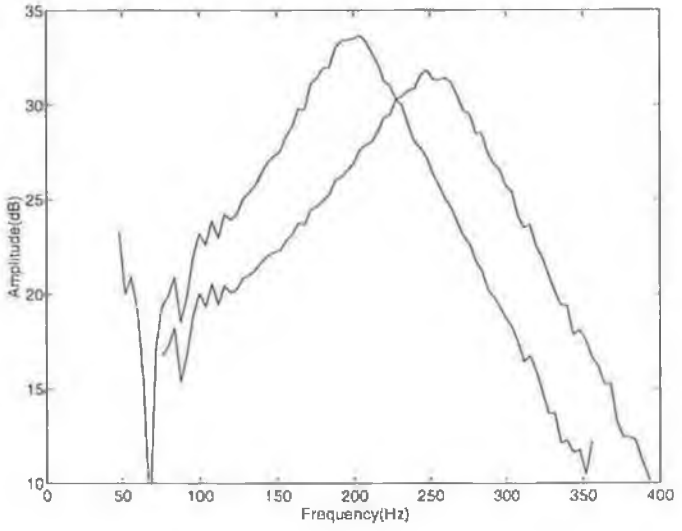


Figure C3: Corrected (higher frequency) and uncorrected resonance curves for subject sc1

In general the correction has the effect of raising the first resonance in frequency, as shown in figure C3 above. This is expected as the volume of the lip adaptor tube is effectively subtracted from the volume of the vocal tract.