Chapter 1 INTRODUCTION

1.1 PROBLEM STATEMENT

A cochlear implant is a prosthetic device that can provide severe-to-profoundly deaf individuals with partially restored hearing (Loizou 1998). It emulates the function of a normal cochlea through combined functioning of externally situated electronics and an electrode array surgically implanted into the cochlea (Figure 1. 1). The electrode array of the Nucleus 24 Contour system consists of 22 half-banded electrodes on a precurved silicone carrier. Electrodes face the modiolus of the cochlea, where the spiral ganglion cell bodies are situated.

Incoming sound is picked up by a directional microphone, from where it is passed to a speech processor that digitises the analogue sound. These signals are continually streamed to the implanted receiver-stimulator by modulating a high-frequency carrier. The stimulator decodes the modulated signal, resulting in activation of appropriate electrodes on the array. Activation of electrodes in turn results in stimulation of the auditory nerve.

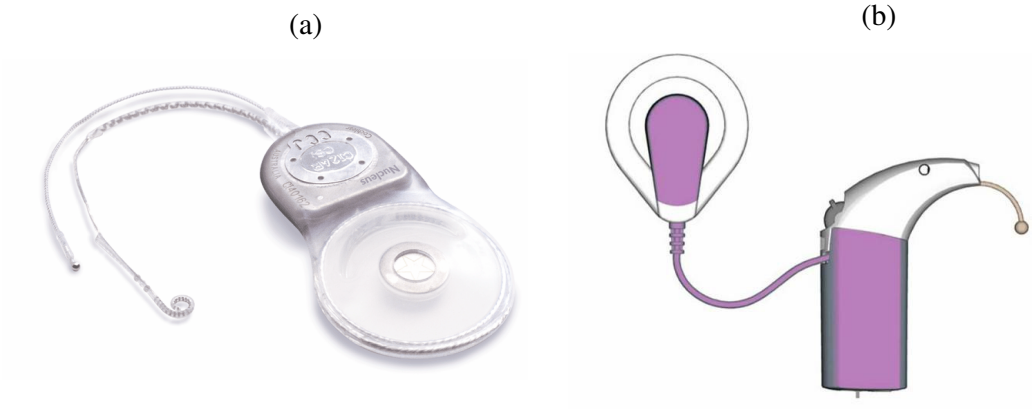

Figure 1.1 (a) The Nucleus 24 Contour cochlear implant, showing the electrode array on the silicone carrier and the subcutaneous stimulator. (b) The Nucleus Esprit 3G speech processor with directional microphone, as well as the external receiver. These components represent the internal and external parts of the implant system.

The speech coding strategies implemented in speech processors aim to stimulate the auditory nerve in a way similar to that of a normal working cochlea by modelling the way the cochlea processes sound. In a normal functioning cochlea, the sound pressure wave is transmitted via the outer ear and vibrates the eardrum. The eardrum and the three middle-ear bones in turn vibrate the oval window, matching the impedance of the cochlea in the way the movement is reduced while increasing the pressure. The vibration of the oval window sets the perilimph fluid inside the cochlea's canals in motion, causing pressure differences between the scala vestibule and scala tympani. These pressure differences cause the basilar membrane, that divides these two scala (if the scala media is ignored), to deflect. This initial deflection, combined with the changing pressure differences initiates a wave travelling from the base of the cochlea towards its apex. Due to the mechanics of the basilar membrane, the amplitude of the travelling wave increase (for a single pure tone) until it reaches a point of maximum deflection before it dies away. The growth of the travelling wave is also accompanied with a marked slowing down of the wave as it approaches the point of maximal deflection. As the basilar membrane deflects from its rest position, tiny hair-cells on the membrane cause the neurons of the auditory nerve in its vicinity to fire, causing a hearing sensation to the person.

At present, commercially available speech coding strategies for cochlear implant speech processors, for example advanced combination encoder (ACE) (Advanced Combination Encoders) and continuous interleaved sampling (CIS) strategies are based mainly on the tonotopical arrangement of the cochlea, i.e. the link between a specific frequency and the point of maximal deflection in the cochlea. For both these strategies, incoming sound is separated into a number of frequency bands and then presented to the auditory nerve via the implant.

In the continuous interleaved sampling (CIS) strategy, the outputs of the filters are sampled at a high rate and presented to the auditory nerve by means of sequential pulses, each filter or frequency band corresponding to an electrode. If the word "*choice*" is processed using the continuous interleaved sampling (CIS) speech processing strategy, the stimuli can be shown in an electrodogram (Figure 1.2), where the vertical axis represents the electrode position (corresponding to the specific filter's frequency) and on the horizontal axis, the time progression. It can be seen from Figure 1.2 how the intensity of the stimulation changes as the consonants are followed by vowels and again by the consonants.

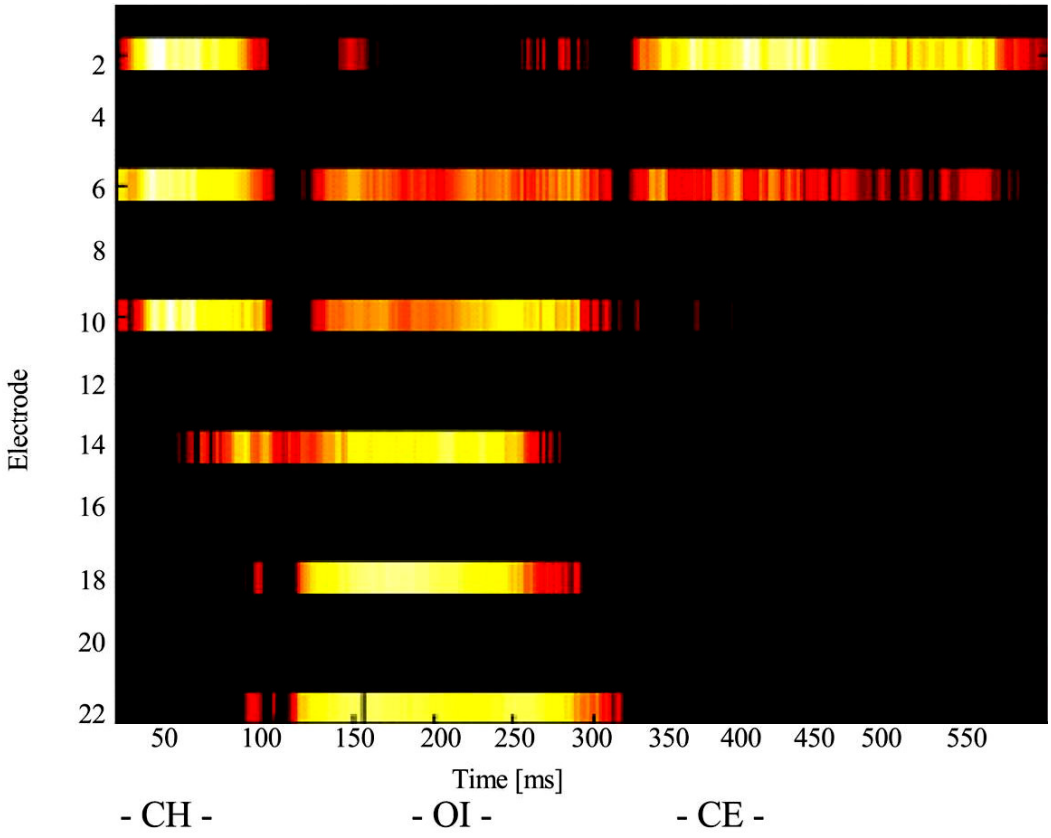

Figure 1.2 Electrodogram of the word 'Choice' processed with continuous interleaved sampling (CIS) speech processing strategy. The colours indicate the amplitude of stimulation on the specific electrode at a specific time – the brighter the colour, the higher the amplitude.

In the advanced combination encoder (ACE) speech coding strategy more channels are used for stimulation, but only the frequency bands or filter outputs with the most energy are presented to the auditory nerve for each time sample. The resultant stimuli pattern is a 'roving'-type where only a limited number of electrodes (typically more than eight) are stimulated during each time cycle, despite being selected from a large number of electrodes (see Figure 1.3). Comparing the electrodogram generated by processing the word "*choice*" with the advanced combination encoder (ACE) strategy to the one from continuous interleaved sampling (CIS) processing, it can be seen that advanced combination encoder (ACE) has an improved frequency resolution, while continuous interleaved sampling (CIS) has a finer temporal resolution due to its higher stimulation rate.

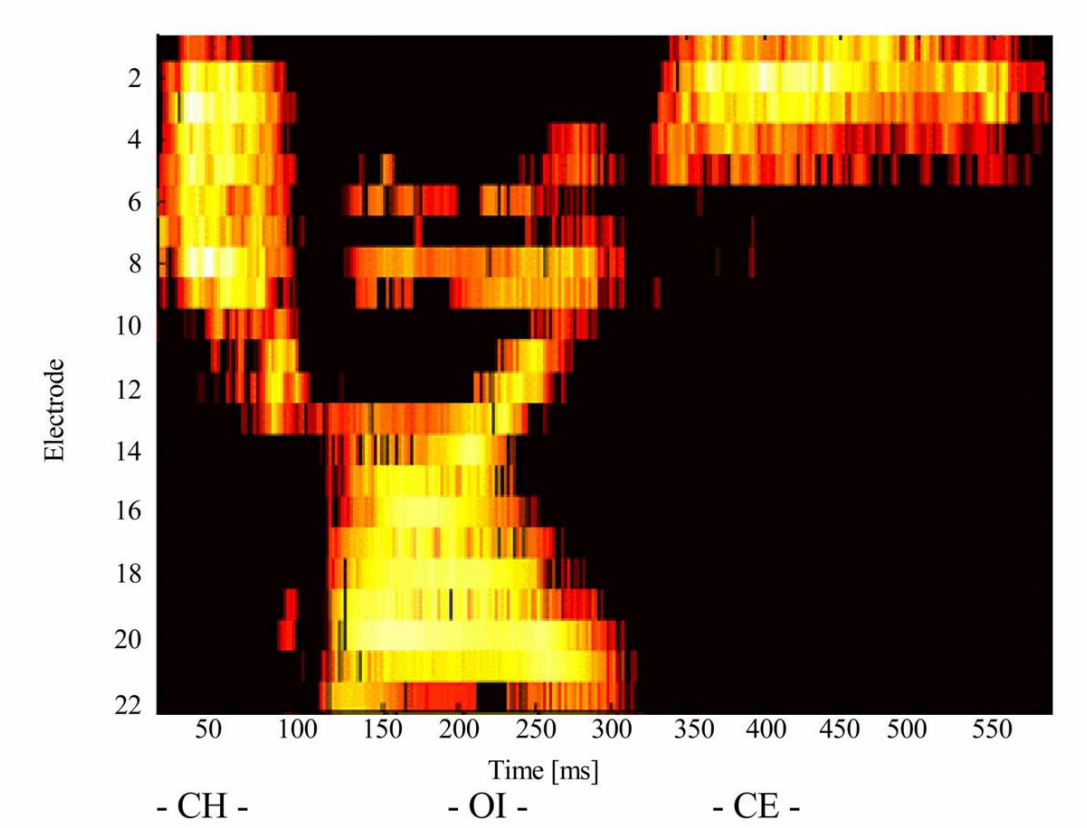

Figure 1.3 Electrodogram of the word 'Choice' processed in advanced combination encoder (ACE). The colour scheme is the same as in Figure 2

However, both these speech processing strategies are limited by the width and overlap of their filters in terms of absolute frequency discrimination, i.e. they lack fine temporal structure of the incoming audio signal. Spectral resolution in current implementations of these strategies is negatively influenced by filters that are more than 100 Hz wide (at -10 dB from the peak response), as shown in Figure 1.4. There is some overlap of the filter frequency responses, resulting in stimulation on two to six electrodes as a pure tone is sweeped from 100 Hz to 1 kHz, with frequency resolution at the low frequencies being affected most. To discriminate between two pure tones that are closely spaced in the frequency domain, the absolute loudness of the two stimuli should be compared on two adjacent electrodes.

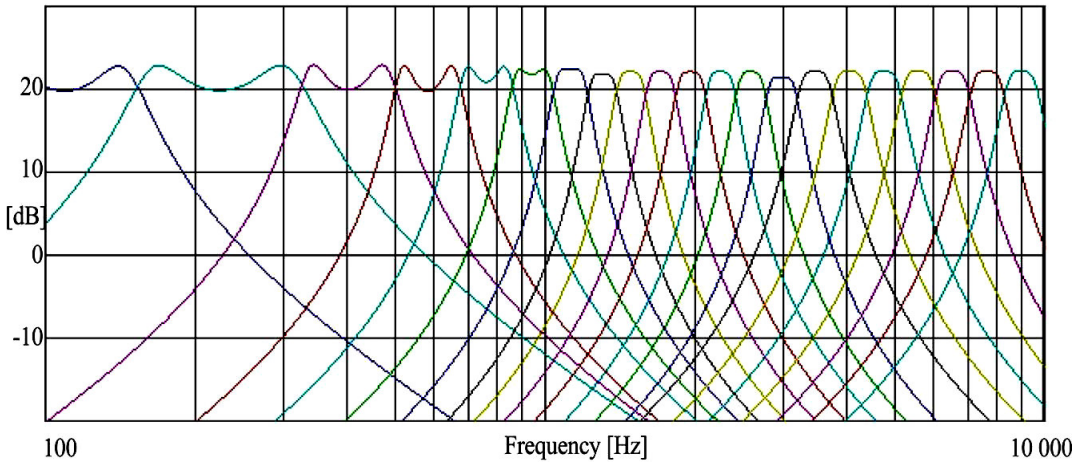

Figure 1.4 Speech processor filterbank frequency responses.

When using only the tonotopisity of the cochlea, the fine time structure of the travelling wave (Von Bekesy, 1947) on the basilar membrane is ignored. The travelling wave caused by pressure differences between the scala tympani and scala vestibuli when the oval window is displaced by the stapes, always starts at the base of the cochlea and moves toward the apex. It increases monotonically in amplitude while travelling towards the apex

until a place of maximum deflection, associated with a specific, characteristic frequency for a pure tone input (Von Bekesy, 1947; Rhode, 1971; Lighthill, 1981; Duifhuis, 1988; Ruggero, 1994), is reached and then rapidly dies away. Although the displacement amplitude of the travelling wave peaks as it reaches this point, the wave velocity slows dramatically, i.e. a rapid increase in phase delay, and its energy is dissipated easily with even the smallest amount of damping (Lighthill, 1981). It is thought possible that the temporal pattern of the basilar membrane (i.e. the synchronous firing of groups of neurones in phase with the negative deflection of the basilar membrane) and the slowing down of the travelling wave (i.e. asymptotic approach of the characteristic frequency over a relatively long time period) may assist a normal hearing individual with the detection of small pitch differences and doing so even in competing in noise. Currently implemented cochlear implant speech processing strategies assume that only the positions of the peaks of the travelling wave envelope carry information, ignoring the existence of travelling waves. The gradual growing and dying away of the travelling wave on the basilar membrane and the rapid increase in phase-shift as the wave approaches the point of maximal deflection is therefore not incorporated in current speech processing strategies such as continuous interleaved sampling (CIS) or advanced combination encoder (ACE).

Users of current speech coding strategies are limited in pitch discrimination ability by the frequency bandwidth assigned to each electrode, which in turn is a function of the number of electrodes implanted. Theoretically this limits a user's discrimination between two adjacent electrodes to a frequency resolution of roughly 100 Hz steps at 100 Hz, quite removed from the almost 1 Hz frequency resolution of a normal hearing individual around 100 Hz.

In this study, the behaviour along the full length of the basilar membrane will be investigated in the time domain, i.e. the deflection along the whole membrane for any point in time, in order to evaluate the relevance of the travelling wave in coding sound in a cochlear implant system. The additional information acquired by emulating the motion of the basilar membrane, will be transmitted to the recipient in the electrical stimulus patterns, to assess whether it provides recipients of cochlear implants with better pitch discrimination.

1.2 RESEARCH QUESTION AND HYPOTHESIS

The primary research question addressed by this study is:

"Can spectral information presented to cochlear implant recipients be improved by incorporating more information regarding the travelling wave?"

The hypothesis is that *when a hydrodynamic model of the basilar membrane is used to act as encoder for sound, spectral information should be more accurately perceived by cochlear implant users than when using current strategies.* This hypothesis was tested by using such a model as a speech coding strategy. As will be shown, sound was processed using a travelling wave emulating algorithm and presented to recipients in a range of experiments. Pitch discrimination studies, during which this travelling wave encoding strategy was compared to current, commercially available strategies, were performed. Following these, pitch ranking experiments was done to establish the tonal quality of the stimuli. Only pure tones were used as input for the speech processing in this study.

1.3 OBJECTIVES

Objectives for this dissertation were the following.

First, an appropriate model of the basilar membrane was selected from literature. Such a model should closely represent the motion of the basilar membrane. The family of hydrodynamic models, describing the motion of the basilar membrane through differential equations, were closely evaluated. These models describe what is actually happening, i.e. it limits the amount of assumptions made about the basilar membrane motion. The outputs of these models are in good agreement with measured data.

Second, a technique to solve these differential equations in the time domain has to be developed. The solution or output of these equations should be the position of the basilar membrane in discrete time steps.

Thirdly, a technique should be developed to convert the output of these equations to a format that can be used by a cochlear implant speech processor to present desired stimuli to a cochlear implant recipient. The solution will therefore be used as a signal processing strategy for the cochlear implant system, indicating where and when the implant should stimulate the cochlea.

Finally, it is necessary to experimentally determine how effective this travelling wave encoding strategy transmits frequency information to the electrically stimulated auditory system, when compared to existing strategies.

1.4 APPROACH

To achieve the objectives set out in the previous paragraph, the following approach will be taken.

1.4.1 Hydrodynamic model

A suitable model that describes the basilar membrane displacement needs to be found from literature. Such a model should closely represent the physical aspects of the cochlea, as well as the motion of both the cochlear fluid and basilar membrane. The differential equations describing the pressure distribution of the fluid on the basilar membrane, resulting in its movement, need to be solved numerically to predict the position of the basilar membrane for discrete time increments. Being able to process any input signal (not just single-frequency sinusoids) will also make such a model suitable for processing of speech and other complex sounds. A balance between complex, accurate models (e.g. three-dimensional models which include macro- and micromechanics inside the cochlea and active sharpening through the outer hair cells) (Neely & Kim, 1986; Wan & Fan, 1991; Böhnke & Arnold, 1999; De Boer & Nuttall, 1999; Steele & Lim 1999) and simplistic models (e.g. one-dimensional models using transmission line theory) (Peterson & Bogert, 1950), is likely to be two-dimensional hydrodynamic models that include only macromechanics of the cochlea without any active elements, and some third-dimension effects that can easily be translated to two dimensions. An example of such an effect is the cross-section curvature of the basilar membrane during deflection (Allen & Sondhi, 1979), which is similar to a clamped string with maximal deflection in the centre that can be approximated by merely looking at the deflection in the centre of the cross-section.

1.4.2 Implementation of model in software

The chosen model will be implemented in Matlab[1]. Two sets of differential equations were used in the model. The first describes the distribution of the pressure on the basilar membrane due to the movement of the stapes and oval window as well as the basilar membrane itself. The second set of differential equations describes the acceleration of the basilar membrane due to the distributed pressure. Solving these equations numerically, the next position of the basilar membrane can be computed using basilar membrane positions from previous time steps and current pressure distribution on the basilar membrane.

A simplified model of the hair cell transfer function is used to act as a 'rectifier' of the basilar membrane movement, increasing the likelihood of the neurons firing on negative deflections.

Application of this model to a cochlear implant imposes a number of constraints. First, the number of relevant implementation sites along the basilar membrane is much less than the amount needed for accurate simulation – 22 electrodes along the length of the cochlea for the implant used. Furthermore, the current Nucleus cochlear implant limits the total stimulation rate to 14 400 pulses per second (pps), which could be, for example 1 000 pulses per second on fourteen electrodes or 2 400 pulses per second on six electrodes. Because of these restraints, the output of the model is sub-sampled in both space and time to 22 discrete sites along the basilar membrane at a total rate of 14 400 samples per second.

1. Matlab is a mathematical interpreting-type package that excels in manipulating matrices in calculations and has good graphic capabilities to display results. www.mathworks.com

The process employed by advanced combination encoder (ACE[2]) is diagrammatically shown in Figure 1.5. Digitised sound in wave-file format (usually with *.wav extension) is used as input. The sound is processed in time blocks, adjusted by a Hanning window to limit boundary-effects during the Fourier transform before going through a fast Fourier transform filter with 64 linear output bins. The output bins are combined using frequency allocation tables to equal the number of channels available in the recipient's MAP, a set of recipient specific parameters that controls the presentation of sound to the recipient to optimise hearing performance. Of the total number of channels present in the recipient's MAP, only those with the most information or highest levels of filter outputs are chosen. The numbers of channels that are presented to the recipient at each time interval are called the number of 'maxima'. The current level necessary for stimulation is determined by converting the filter output from each of the selected channels to a logarithmic current scale by means of a non-linear transfer function.. This information is then transmitted to the cochlear implant by modulating the radio frequency carrier that also provides power.

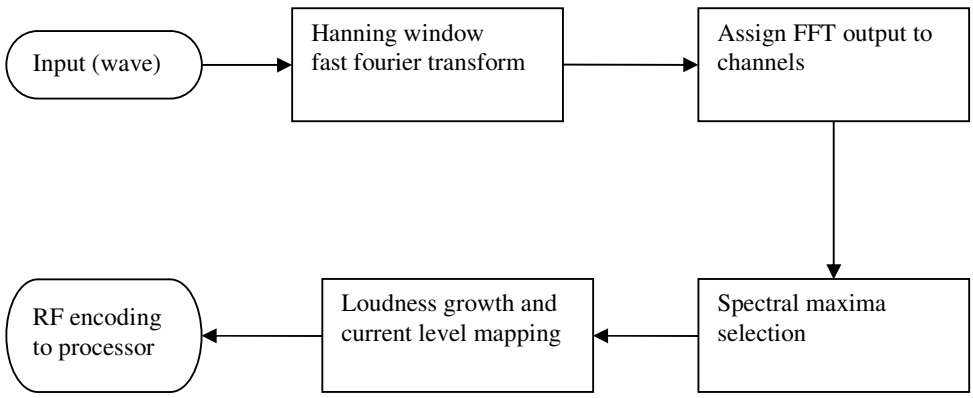

Figure 1.5 Processing steps in advanced combination encoder (ACE) encoding strategy. The input is in wave-sound file format.

2. Advanced combination encoder (ACE) is the default speech coding strategy used with the Nucleus 24 cochlear implant system of Cochlear Ltd., a leading, Australian based cochlear implant manufacturer. www.cochlear.com.au

The travelling wave encoding strategy retains most of the processes used for the advanced combination encoder (ACE) encoding strategy in order to ease implementation and comparison of their respective results. Digitised pure tones in wave-sound file format are also used as input, which the hydrodynamic model block uses to determine the displacement of the basilar membrane for each sampled time increment (Figure 1.6). To comply with the restrictions of the current Nucleus cochlear implant system, the basilar membrane displacement is sub-sampled in both time and space, with sub-sampling in space done on 22, linearly spaced positions along the basilar membrane. Electrodes to be stimulated were selected similarly to the maxima selection in the advanced combination encoder (ACE) strategy, i.e. the highest values of the basilar membrane displacement was selected for stimulation of the corresponding electrode. The loudness growth transfer function that computes the current levels corresponding to basilar membrane displacement was changed to a linear function, instead of the non-linear one used in the advanced combination encoder (ACE) strategy. This was done to ensure that the basilar membrane displacement is analysed without distortion of non-linear transfer functions. The RF coding and transmission of the individual stimuli were kept the same as in the advanced combination encoder (ACE) strategy.

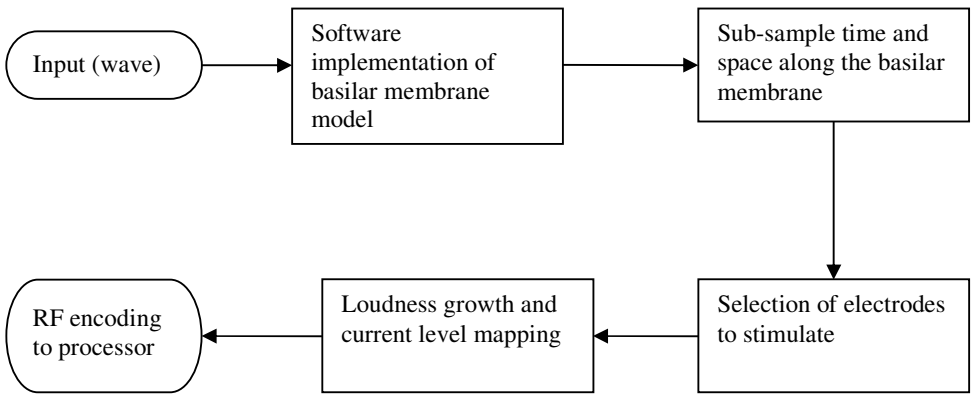

Figure 1.6 Processing steps of the proposed encoding strategy.

1.4.3 Evaluating the model (experimental work)

To evaluate the implementation of the model as a possible speech processing strategy, a first step is to compare listeners' pitch discrimination abilities for pure tones processed by the advanced combination encoder (ACE) strategy (Figure 1.5) and the travelling wave encoding strategy (Figure 1.6) respectively. There is a suggested link between increased pitch discrimination ability and improved speech perception (Moore, 1986; Hanekom & Shannon, 1996; Henry, McKay *et. al.*, 2000). By using simple stimuli in an experimental setup where sound can be pre-processed, the potential usefulness of a new strategy can be evaluated without the need to implement the strategy in real time within the constraints of the speech processor.

To ensure that discrimination is not linked to loudness, an adaptive loudness balancing procedure will be performed to eliminate loudness cues from the different stimuli. It will be followed by a discrimination experiment using 21 pure tones in 1 Hz increments, around 100 Hz, to establish the ability of the recipient to distinguish between each of the 21 tones. A similar experiment will be done with 21 pure tones in 10 Hz increments, around 1 kHz. The output of this experiment will be displayed in a confusion matrix that is expected to have a triangular form, with good discrimination for tones far apart and poorer discrimination for tones close together.

If sufficient discrimination is achieved with the travelling wave encoding strategy, a pitch-ranking experiment will be done to establish whether the new sounds only differ from each other or whether they can also be assigned a rising tonal quality linked to the rising frequency of the inset stimuli.

1.5 CONTRIBUTION

This work contributes to current knowledge by implementing a more complex basilar membrane model as speech processing algorithm in cochlear implant speech processors. The model incorporates the travelling wave on the basilar membrane and adds both temporal pitch, by way of the periodic deflection of the basilar membrane, and the deceleration effect of the wave as it approaches its characteristic frequency, to the information presented to the cochlear implant recipient.

It also provides a basic model from which more complex models of the outer, middle and inner ear (and hair cell function) can be implemented during future research. The modular approach of the Nucleus Matlab toolbox not only makes the additional processing steps created for current experiments ideal building blocks for future research, but will also be easily compatible with additional building which may be added to the proposed speech coding strategy.

1.6 OVERVIEW

In the second chapter, a detailed literature study will be presented, explaining the background of the problems experienced with current speech coding strategies implemented in cochlear implant speech processors. Models describing the functioning of the basilar membrane will be discussed, as well as their possible applications for and effects on frequency and speech detection. Possible solutions will also be discussed.

In the third chapter, the different hydrodynamic models will be discussed, considering their respective strengths and weaknesses. One-, two- and three-dimensional models with solutions in both time- and frequency-domain will be explored and discussed. One such two-dimensional model with solutions in the time-domain will be chosen from these models and implemented in Matlab software. Problems encountered during this process, as well as solutions or alternatives, will also be discussed in this chapter.

The fourth chapter will focus on the integration of the above mentioned model into the Nucleus Implant Communicator (NIC) toolbox in Matlab, to provide a fast-prototyping of this model and comparison with current, commercial strategies. Cochlear Ltd.'s NIC software will be used as platform to do a rapid prototype evaluation of sound processing using this model. NIC acts as a software interface between the programming system and speech processor. It consists of a dynamic linked library (dll) and a toolbox written for use with Matlab (Nucleus Matlab Toolbox). Using the toolbox functions in Matlab, almost any possible stimulation pattern can be presented to a recipient. Detailed discussion of how the toolbox was used will be used and changes made to the model to integrate it with the toolbox will be discussed in Chapter 3.

The fifth chapter will discuss the experimental procedure followed. It will mention the case histories of the recipients that participated in the study and will focus specifically on electrodograms and captured radio-frequency data that show the stimulus pattern and levels of stimulation. The experimental setup of each experiment will be discussed and results will be shown.

The sixth chapter will discuss the results obtained during the experimental sessions. Results obtained from current speech coding strategies will be compared to those from the evaluated strategy and discussed in terms of the hydrodynamic model. The effectiveness of using the discussed model as a basis for speech coding strategies will be evaluated, followed by some recommendations for improvements and further investigations of using the same building blocks.

In the final chapter, the dissertation will be analysed and discussed in terms of the hypothesis, objectives and problem stated in the first chapter.

Chapter 2 LITERATURE STUDY

2.1 INTRODUCTION

In this chapter the background of current, commercial speech coding strategies and some of the problems still experienced with these are discussed. Various, more complex possibilities of how the cochlea processes sound is also discussed and compared with one another.

2.2 BACKGROUND LITERATURE STUDY

The concept of the travelling wave on the basilar membrane has long been known (Von Bekesy, 1947) and is commonly accepted (Allen, 1977; Shamma, 1985a; Duifhuis, 1988; Prosen & Moody, 1991; Ruggero, 1994; Uppenkamp, Fobel & Patterson., 2001; Moore, Alcatrana & Glasberg, 2002), yet the models of the cochlear mechanics implemented in current, commercial cochlear implant speech coding strategies only include the points of maximal deflection on the basilar membrane (Loizou, 1998). Considering only the tonotopicity of the cochlea simplifies speech processing to only filtering of incoming sound, followed by e.g. maxima selection and current level mapping when using advanced combination encoder (ACE). This basic model of how the cochlea processes sound led to the development of signal processors that used symmetrical filter-banks to decompose the input signal into its spectral components (Loizou, Dorman and Fitzke, 2000).

Loizou (1998) gave a short overview of the history of cochlear implants and existing speech processing strategies. Speech perception performance of cochlear implant recipients improved over the years, with many recipients being able to understand speech without lip-reading and some even able to use telephones (Loizou,1998). Speech

production skills are closely related to speech perception and Loizou reported gradual improvements over time up to four years after the start of using the cochlear implant.

One of the improvements of continuous interleaved sampling (CIS) and advanced combination encoder (ACE) over previous speech processing strategies is the ability to encode more temporal information into the stimuli presented to the cochlea, i.e. representing rapid envelope variations (Loizou, 1998). The travelling wave encoding strategy continues along this line by not only considering the envelope variations, but also looking the travelling wave on the basilar membrane, in an attempt to incorporate it in the stimuli presented to the cochlear implant recipient.

Other parameters of how the cochlea processes sound have also been reported. Von Bekesy (1947) suggested that phase measurements may be better in determining the pattern of vibration (i.e. position and pitch) than considering the amplitude envelope alone. Shamma (1985a) has shown that the distribution of the relative phases of synchronised activity in the auditory nerve is an important indication of spectral parameters. He has shown in particular that, for regions basal to the characteristic frequency point on the basilar membrane, the relative phase shifts are small (fast propagating travelling wave) in comparison to those at the characteristic frequency point as the travelling wave decelerates. These rapid phase shifts can be seen in the auditory nerve as steep and localised spatial discontinuities along the tonotopical axis of fibres that otherwise fire in a relatively uniform instantaneous pattern if the frequency is lower than the characteristic frequency point. This is visible for the characteristic frequency locations that correspond to low frequencies (less than or equal to 1.5 – 2 k Hz for cats). Shamma (1985b) further argues that lateral inhibitory networks contribute to processing of these localised spatial discontinuities. In his defence of the travelling wave phenomenon in the cochlea, Ruggero

(1994) also uses the experimentally measured phase delays associated with points close to the characteristic frequency as proof of the existence of the travelling wave. It can be concluded from the above that the change in phase shifts and deceleration of the travelling wave close to the characteristic frequency carry some useful information regarding the pitch of incoming sound. Using a model of the basilar membrane that can show such deceleration of the travelling wave should therefore be able to convey this information to the cochlear implant user.

An implementation that used phase information of the incoming sound to time the stimuli of the cochlear implant was developed by Hartmann, Tönder and Klinke (1998), but it failed to reflect the changing phase delays seen in the cochlea as the travelling wave travelled on the basilar membrane. In order to simplify calculation, it ignored the relative phase shifts and in particular the sharp discontinuity in phase shift/delay, seen close to the characteristic frequency. This implementation was not clinically evaluated.

In the normal cochlea, for low stimulation rates, the neural firing pattern locks on to the phase of the travelling wave and is most likely to stimulate when the basilar membrane is moving towards the scala vestibuli, with almost no stimuli when moving towards the scala tympani (Rose, Hind, *et. al.,* 1971). A high degree of phase locking exists up to 1 kHz, after which it declines to approach zero at 5 kHz. This dependence on stimulus frequency can be attributed to the capacitance of the basal walls of the inner hair cells (Russel & Sellick, 1978). Maximum firing rate of auditory fibres appear to be affected by the saturation of the input/output function of the inner hair cells (for frequencies lower than the characteristic frequency), as well as the basilar membrane non-linearity at frequencies higher than the characteristic frequency, resulting in the travelling wave not propagating beyond the characteristic frequency position (Pattuzzi & Sellick, 1983). Moore (1986)

suggested, in general agreement with the above, that information in the phase locking of neural pulses up to 5 kHz is essential for effective frequency discrimination. He further suggests that both phase locking information and place information are used for complex stimuli. The relevance of these outcomes to the proposed study will be discussed extensively in Chapter 6, but it implies that, in the low frequency range, the information contained within the periodic displacement of the basilar membrane can convey additional pitch information to the cochlear implant recipient.

In current speech coding strategies, a non-linear transfer function is used when processing sound (e.g. loudness growth function in Figure 1.6). Such an additional transfer function should not be necessary for the travelling wave encoding strategy if the non-linearity of the basilar membrane displacement is incorporated in the hydrodynamic model. The non-linear response of the basilar membrane has been attributed to the non-linear transfer function of the stereocillia deflection to the receptor potential of the outer hair cells (Preyer & Gummer,1996). Others have suggested that the mechanical non-linearity of the Dieters cells is a more relevant cause for the non-linear basilar membrane motion (Böhnke & Arnold, 1998). Sellick, Pattuzzi and Johnstone (1982) stated that the cochlear mechanics are actually sharply tuned in comparison to neural tuning, which can result in the sharp tuning seen in inner hair cells and auditory nerve fibres. In general, it can be stated that the stimulation of the auditory nerve follows the low frequency movement of the basilar membrane. The relatively broad sensitive region around 1 kHz that is independent of the position in the cochlea might be linked to the change in viscous coupling from velocity-related to displacement-related coupling (Pickles, 1986).

2.3 SUMMARY

In conclusion, current cochlear implant speech processing strategies provide users with good speech discrimination, but are as yet unable to convey all the auditory information needed for good music appreciation and discrimination of closely spaced pitch differences at low frequencies. It is argued that this inability is intrinsic to the current model of the cochlea and basilar membrane underlying current commercial speech coding strategies. This model reduces the cochlea and basilar membrane to a simple tonotopic function that stimulates auditory nerves only at the point corresponding to the characteristic frequency and all across the cochlea at roughly the same time. However, development of speech processing strategies that convey fine temporal structure of the envelope (e.g. advanced combination encoder (ACE) and continuous interleaved sampling (CIS)) has significantly improved recipients' speech perception performance. This study aims to further improve performance by not only providing additional information regarding fine temporal structure, but by also re-evaluating the basilar membrane model that is used. The concept of a travelling wave on the basilar membrane is well accepted when considering the functioning of a normal hearing individual's cochlea. The implementation of hydrodynamic models in cochlear implant speech processing strategies has until now, however, been prevented by the sheer computational power needed to solve such models – even with the travelling wave encoding strategy sound currently needs to be processed prior to stimulation because of the amount of time required for information processing. Implementation of this model provided additional information about the phase and progressive change of phase delay of stimuli, period of low frequency sounds that enables the recipient to discriminate pitch because of phase locking of the auditory nerve and the intrinsic non-linear amplitude response of the basilar membrane.

Chapter 3 HYDRODYNAMIC MODEL AND IMPLEMENTATION

3.1 INTRODUCTION

Following the discussion in the previous chapter, the different hydro-dynamic models will be discussed in this chapter, considering their various strengths and weaknesses. One-, two- and three-dimensional models with solutions in both the time- and frequency-domains will be explored and discussed. From these models one will be chosen and implemented in Matlab software. Problems encountered during this process (with their solutions or alternatives) will also be discussed.

3.2 BACKGROUND

In order to overcome the constraints of current models, mathematically modelling the fluid mechanics of the cochlea was investigated. As computing power and modelling methodology improved, increasingly complex models were developed. The earliest of these were one-dimensional (transmission line) models, which showed only qualitative correlation with experimental values (Peterson & Bogert, 1950) - specifically a broad area of maximal deflection around the characteristic frequency that was not observed in later measurements (Rhode, 1971). Two-dimensional models, incorporating two-dimensional flow of fluid in the scala and output in the frequency domain were subsequently developed (Allen, 1977; Sondhi, 1978). These models were refined to consider longitudinal stiffness but more pertinent to the aim of the current study, to also provide output in the time domain (Allen & Sondhi, 1979). Refinements of such two-dimensional models included the addition of the effect of viscous fluid on the whole model (Wan & Fan, 1991). Subsequent three-dimensional models analysed the effect of a viscous fluid (Steele & Taber, 1979b).

Johnstone, Pattuzzi and Yates (1986) discussed some of the experimental evidence pointing to possible 'feed forward' elements that contribute to the non-linear amplitude response of the basilar membrane. Such response behaviour is characterised by high sensitivity and sharp spectral tuning at low amplitudes which decreases with increasing amplitude levels. Although one of the earlier models that included active elements (Neely & Kim, 1986) was linear and only in one dimension, it included micromechanics, considering longitudinal stiffness and using eardrum pressure as input. As computing power increased, some older two-dimensional models were greatly refined to include the micromechanics of the Organ of Corti (Allen & Neely, 1992) as well as active elements. More recently, active elements were also included in three-dimensional models (Steele & Lim, 1999; De Boer & Nuttall, 1999), all implemented in the frequency domain.

The use of finite difference and finite element methods further assisted the solution of this complex hydrodynamic system. Finite difference methods were first used in two dimensions, since they required less computing power and were easier to implement directly from the differential equations (Allen & Sondhi, 1979; Steele & Taber, 1979a; and Neely, 1981). Implementation in three dimensions followed later (Steele & Taber, 1979b). Similarly, finite elements were also used first in two-dimensional solutions (Janssen, Segal & Viergever, 1978), followed by later three-dimensional solutions which included the micromechanics (Organ of Corti) (Böhnke & Arnold,l 1998) as well as fluid – structure couplings (Böhnke & Arnold, 1999).

For purposes of this study, the model of Allen and Sondhi (1979) was most useful, since it is the most advanced existing model that describes the position of the basilar membrane in the time domain. Since it is aimed to link the basilar membrane position for each sampled

point in time as a reference for stimulation, the output has to be in the time domain. This solution also includes some non-linear basilar membrane properties that are best modelled in the time domain, since working in the frequency domain has various assumptions that rely on linear superposition.

3.3 IMPLEMENTATION

Allen and Sondhi (1979) describe the fluid motion and that of the basilar membrane motion separately by using Green's theorem to relate the distributed pressure across the basilar membrane to its acceleration. Together with the plate equation that describes the basilar membrane motion resulting from this acceleration, these two equations completely specify the basilar membrane motion resulting from movement of the oval window. This model assumes the cochlear fluid to be incompressible, does not account for coiling and also excludes lateral fluid motion (movement in the third dimension). Some non-linear basilar membrane properties, for example damping, are included.

Allen (1977) extended these assumptions so that basilar membrane velocity and scala pressure are related through by the integral equation

$$p(x,t) = \rho \int_0^L G(x,x') \ddot{\xi}(x',t) dx' + (L-x)\rho \dot{u}(t) \qquad \textbf{(3.1)}$$

where $p(x,t)$ represents the scala pressure, ρ the cochlear fluid density, L the length of the cochlea, $\xi(x,t)$ the basilar membrane displacement, x the position along the basilar membrane and $u(t)$ the stapes velocity. G represents Green's function. Green's function has been derived previously (Allen, 1977; Sondhi, 1978) but is in principle the solution to Laplace's equation for the pressure distribution on the basilar membrane, given the acceleration at any point along the membrane. Allen and Sondhi (1979) unfolded the

cochlea and transformed it into a periodic extension with period 4L, resulting in a 'real-odd harmonic' symmetry that can be expanded as a Fourier series, the following equations were mostly taken from Allen and Sondhi (1979). This symmetry is so called because the Fourier series expansion of such a periodic function consists only of odd-harmonic cosine terms.

$$2p((x)) = \rho\int_{-2L}^{2L} F((x-x'))\ddot{\xi}((x'))dx' + 2((L-|x|))\rho\dot{u} \qquad \textbf{(3.2)}$$

This then leads to where the replacement $G(x,x') = F(|x-x'|) + F(|x+x'|)$ has been made from Equation 3.1 and F has real-odd harmonic symmetry. The double brackets denote the function as being periodic with period 4L.

Equation 3.2 summarises the fluid coupling inside the cochlea and the resultant pressure on the basilar membrane. The second part of Allen and Sondhi's (1979) hydrodynamic model describes the basilar membrane displacement, $\xi(x,t)$, in terms of all the forces acting on it. If the basilar membrane is modelled as a homogeneous, anisotropic plate, the equation for the plate's displacement is

$$\mathrm{q(x,z,t)} = D_x\frac{\partial^4\xi}{\partial x^4} - 2(D_x D_z)^{1/2}\left(\frac{\pi}{\mathrm{W}(x)}\right)^2\frac{\partial^2\xi}{\partial x^2} + D_z\left(\frac{\pi}{\mathrm{W}(x)}\right)^4\xi \qquad \textbf{(3.3)}$$

where q is the load on the membrane and the parameters D_x and D_y are functions of the internal structure, thickness and modulus of elasticity of the basilar membrane. These parameters are assumed to be constants but not necessary equal (homogeneous and anisotropic).

The load on the basilar membrane consists of the inertia of the basilar membrane $m\ddot{\xi}$, the viscous damping $R(x)\dot{\xi}$ and the trans-basilar membrane pressure, $2p$. When combining

Equations 3.2, 3.3 and these loads on the basilar membrane, the basilar membrane motion is given by

$$D\xi + R(x)\dot{\xi} + m\ddot{\xi} = -\rho F \otimes \ddot{\xi} - 2(L - |x|)\rho\dot{u} \qquad \textbf{(3.4)}$$

where

$$D\xi = D_x \frac{\partial^4 \xi}{\partial x^4} - 2(D_x D_z)^{1/2}\left(\frac{\pi}{W(x)}\right)^2 \frac{\partial^2 \xi}{\partial x^2} + D_z\left(\frac{\pi}{W(x)}\right)^4 \xi \qquad \textbf{(3.5)}$$

The symbol $\otimes$ denotes the circular convolution $\rho\int_{-2L}^{2L} F(x-x')\dot{v}(x')dx'$ with each term having a period of 4L, allowing it to be expanded into a Fourier series. R*(x)* denotes the damping, ξ the basilar membrane displacement (cm), $\dot{\xi}$ the basilar membrane velocity (mm/s) and $\ddot{\xi}$ the basilar membrane acceleration in (cm/s^2). ρ is the cochlear fluid density, L the length of the unrolled cochlea (mm), x the position along the basilar membrane (mm) and $\dot{u}$ the velocity of the stapes movement (mm/s). *Dx* and *Dz* are the longitudinal and transversal stiffness of the basilar membrane respectively, while W*(x)* describes the basilar membrane width as a function of *x*.

This equation can be solved as an initial value problem in time, with boundary values in space. The boundary problem is solved using Fourier transforms, while the initial value problem is solved recursively, starting with the basilar membrane's initial displacement and velocity. Equations 3.4 and 3.5 can be solved numerically if the space and time variables are discretisised as follows:

$$\ddot{\xi}(t) = F_s^2(\xi_n - 2\xi_{n-1} + \xi_{n-2}) \qquad \textbf{(3.6)}$$

and

$$\dot{\xi}(t) = F_s\left(\xi_{n-1} - \xi_{n-2}\right) \tag{3.7}$$

where n is the discrete time interval and F_s is the sampling frequency. The circular convolution in Equation 3.4 can also be solved in the discrete time domain when written as $\frac{\rho L F_s^2}{M}\sum_{m=1}^{4M} F_{k-m}\left(\xi_{m,n} - 2\xi_{m,n-1} + \xi_{m,n-2}\right)$. M represents the number of spatial samples along the basilar membrane, k the discrete spatial intervals and both F_{k-m} and ξ_m is periodic with a period of 4M. A sampled version of F_{k-m} is given by Allen and Sondhi (1979),

$$F_k = \begin{cases} \dfrac{L}{2H} - \dfrac{1}{\pi}\left\{\ln\left(\dfrac{\pi L}{2MH}\right) - 1\right\}, k = 0 \\ \dfrac{L}{2H}\left(1 - \dfrac{k}{M}\right) - \dfrac{1}{\pi}\ln\left(1 - e^{\frac{-\pi k L}{HM}}\right), k \neq 0 \end{cases} \tag{3.8}$$

where H is the height of the scala.

If the terms containing $\ddot{\xi}$ in the discretisised Equations 3.4 and 3.5, are moved to the left, it results in

$$\frac{\rho L F_s^2}{M}\sum_{m=1}^{4M} F_{k-m}\left(\xi_{m,n} - 2\xi_{m,n-1} + \xi_{m,n-2}\right) + mF_s^2\left(\xi_n - 2\xi_{n-1} + \xi_{n-2}\right) = b(x, t) \tag{3.9}$$

where

$$b(x,t) = -D\xi - R(x)F_s\left(\xi_{n-1} - \xi_{n-2}\right) - 2\left(1 - \frac{|k|}{M}\right)L\rho F_s\left(u_{n-1} - u_{n-2}\right) \tag{3.10}$$

To solve for ξ, the left of Equation 3.9 needs to be changed in such a way that $\ddot{\xi}$ can be isolated. Allen and Sondhi (1979) solved this by using the Fourier transform, thereby reducing the convolution to a multiplication in the frequency domain.

In order to do so, they defined

$$Q(x) = \rho F(x) + \left(\frac{m}{2}\right)\left(\delta(x) - \delta(2L - x)\right) \tag{3.11}$$

where $\delta(x)$ is the Dirac δ function, with the factor of 0.5 and the second $\delta(x)$ -term necessary to give Q(x) real-odd harmonic symmetry. Equation 3.9 and Equation 3.10 can now be solved, using the Fourier transform.

$$F_s^2\left(\xi_{\mathrm{n}} - 2\xi_{\mathrm{n-1}} + \xi_{\mathrm{n-2}}\right) = \mathbf{F}^{-1}\left(\frac{\mathbf{F}\{b(x,t)\}}{\mathbf{F}\{Q(x)\}}\right) \tag{3.12}$$

Computing $\mathbf{F}\{Q(x)\}$ once therefore computes the model. The following steps are subsequently iterated until all the input samples are processed.

- Compute $b(x,t)$ in Equation (3.10) using previously computed values of ξ.
- Compute the fast Fourier transform of this result.
- Divide by the previously computed $F(Q_m)$.
- The inverse fast Fourier transform will yield acceleration vector $\ddot{\xi}$.
- Using recursive methods, the next displacement vector ξ, is computed.

The resulting estimate of the basilar membrane position is proportional to the probability of the nerves firing at each point, with a higher probability in the case of deflection in the negative direction.

During implementation of the model of Allen and Sondhi (1979) in Matlab, some problems were encountered.

- The spatial fast Fourier transform, when first implemented, caused zeros in the frequency domain. These zero points prevented the division in the third step of the above algorithm. It was later realised that they were created because of the periodic extension of function F prior to computing the fast Fourier transform, knowing that the fast Fourier transform inherently expand any such function. The initial inability to prevent these zeros resulted in seeking alternative methods for computing the basilar membrane movement.

  An alternative method of solving Equations 3.9 and 3.10 required many repetitions to minimise the error function, since the expression could not be solved explicitly. The second-order terms (unknowns) were moved to the left of the equation with the standard and first-order terms (known values of previous calculations) on the right. The method attempting to solve Equations 3.9 and 3.10 without using the fast Fourier transform follows.

  If the terms containing ξ_n are isolated on the left, the following is obtained:

$$\mathrm{m}F_s^2\xi_\mathrm{n} + \frac{\rho \mathrm{L}F_s^2}{\mathrm{M}} \sum_{m=0}^{4\mathrm{M}-1} \mathrm{F}_{\mathrm{k}-\mathrm{m}} \left(\xi_{\mathrm{m,n}}\right) = \mathrm{A}(\xi) \tag{3.13}$$

where

$$\mathrm{A}(\xi) = -\mathrm{D}\xi - \mathrm{R}(x)F_\mathrm{s}\left(\xi_{\mathrm{n-1}} - \xi_{\mathrm{n-2}}\right) - \mathrm{m}F_s^2\left(-2\xi_{\mathrm{n-1}} + \xi_{\mathrm{n-2}}\right)\ldots$$
$$-\frac{\rho \mathrm{L}F_s^2}{\mathrm{M}} \sum_{m=0}^{4\mathrm{M}-1} \mathrm{F}_{\mathrm{k}-\mathrm{m}} \left(-2\xi_{\mathrm{m,n-1}} + \xi_{\mathrm{m,n-2}}\right) - 2\left(1 - \frac{|k|}{\mathrm{M}}\right) \mathrm{L}\rho F_\mathrm{s}\left(u_{n-1} - u_{n-2}\right) \tag{3.14}$$

$D\xi$ is the same as in Equation 3.5.

ξ_n (the new displacement vector of the basilar membrane) can now be solved using numerical methods. $A(\xi)$ can be computed from already known values of ξ (ξ_{n-1} and ξ_{n-2}) and the input variables u_{n-1}, u_{n-2}. If the summation in Equation 3.13 is expanded,

$$mF_s^2\xi_n + \frac{\rho L F_s^2}{M}\begin{bmatrix} F_{1-1}\xi_{n,1} & F_{2-1}\xi_{n,1} & \cdots & F_{M-1}\xi_{n,1} \\ F_{1-2}\xi_{n,2} & F_{2-2}\xi_{n,2} & \cdots & F_{M-2}\xi_{n,2} \\ \cdots & \cdots & \cdots & \cdots \\ F_{1-4M}\xi_{n,M} & F_{2-4M}\xi_{n,M} & \cdots & F_{M-4M}\xi_{n,M} \end{bmatrix} = A(\xi) \qquad \textbf{(3.15)}$$

where i in $\xi_{n,i}$ denotes the position along the basilar membrane. If each position along the basilar membrane considered individually (i.e. the columns in Equation 3.15),

$$mF_s^2\xi_{n,i} + \frac{\rho L F_s^2}{M}\left(F_{i-1}\xi_{n,1} + F_{i-2}\xi_{n,2} + F_{i-3}\xi_{n,3} + \ldots + F_{i-4M}\xi_{n,4M}\right) = A(\xi_i) \qquad \textbf{(3.16)}$$

for every i, where

$$A(\xi_i) = -D\xi_i - R(x)F_s\left(\xi_{n-1,i} - \xi_{n-2,i}\right) - mF_s^2\left(-2\xi_{n-1,i} + \xi_{n-2,i}\right) - K_i \ldots \\ -2\left(1 - \frac{|i|}{M}\right)L\rho F_s\left(u_{n-1} - u_{n-2}\right) \qquad \textbf{(3.17)}$$

and

$$K_i = \frac{\rho L F_s^2}{M}\left(F_{i-1}\left(-2\xi_{n-1,1} + \xi_{n-2,1}\right) + \ldots + F_{i-4M}\left(-2\xi_{n-1,4M} + \xi_{n-2,4M}\right)\right) \qquad \textbf{(3.18)}$$

Equation 3.16 can be written in matrix form, so that it can easier be solved in Matlab:

$$m F_s \begin{bmatrix} \xi_{n,1} & \dots & \xi_{n,M} \end{bmatrix} - \frac{L\rho F_s^2}{M} \begin{bmatrix} \xi_{n,1} & \dots & \xi_{n,4M} \end{bmatrix} \begin{bmatrix} F_0 & \dots & F_{M-1} \\ F_{-1} & \dots & F_{M-2} \\ \dots & \dots & \dots \\ F_{1-4M} & \dots & F_{-3M-1} \end{bmatrix} = \begin{bmatrix} a_1 & \dots & a_M \end{bmatrix} \quad \textbf{(3.19)}$$

Since Equation 3.17 can be solved using previous displacement values of the basilar membrane and displacement of the oval window, the right side of Equation 3.19 is known. ξ_n can now be solved by choosing a starting value for ξ_n and computing the left side of Equation 3.19. The vector values on both sides of Equation 3.19 will differ. To solve for the correct value of ξ_n, the difference between the left and right side of Equation 3.19 should be minimised. Repetitions, with appropriate minimising or zero-seek methods resulted in a reasonably accurate guess with an acceptably small error.

- The method described above required excessive computing times, e.g.168 hours to compute the basilar membrane movement during 25 ms of sound. Attempts to shorten the computing time by accepting larger errors or by increasing step sizes in approaching the zero error caused the models to become unstable.
- Zero-seek methods, using the gradient of changes in the error function to reduce the number of iterations without causing instability, did not reduce the time required to calculate the basilar membrane response to an acceptable level. To minimise the error typically required between a few and tens of millions of iterations.
- In order to minimise the above mentioned calculation time, multiresolution analysis was investigated (Vetterli, 1991). The spatial resolution in the basal portion of the cochlea needs to be higher than in the apical portion, in order to have an accurate and stable calculation of basilar membrane movement. Analysing the basilar

membrane in several portions, with varying spatial sampling, helped to minimise the size of the matrices and therefore computational time, without the subsequent restrictions on highest analysed input frequency.

- Functions for finding zero points built into the MATLAB toolboxes also improved calculation time, but often led to instability.

As mentioned above, when using the equation from Allen and Sondhi (1979) without expanding it in the frequency domain, the above problems could be solved. With current MATLAB code, one second of basilar membrane movement due to a sound, takes approximately 90 seconds to compute and store when simulated on a Pentium IV, 1.4 GHz processor, using a Windows 2000 operating system.

3.4 RESULTS AND DISCUSSION

When using similar cochlear parameters, the results obtained are in good agreement with the graphical output of the model used by Allen and Sondhi (1979). The steep roll-off above the characteristic frequency can clearly be seen as the travelling wave slows down to approach characteristic frequency (Figure 3.1).

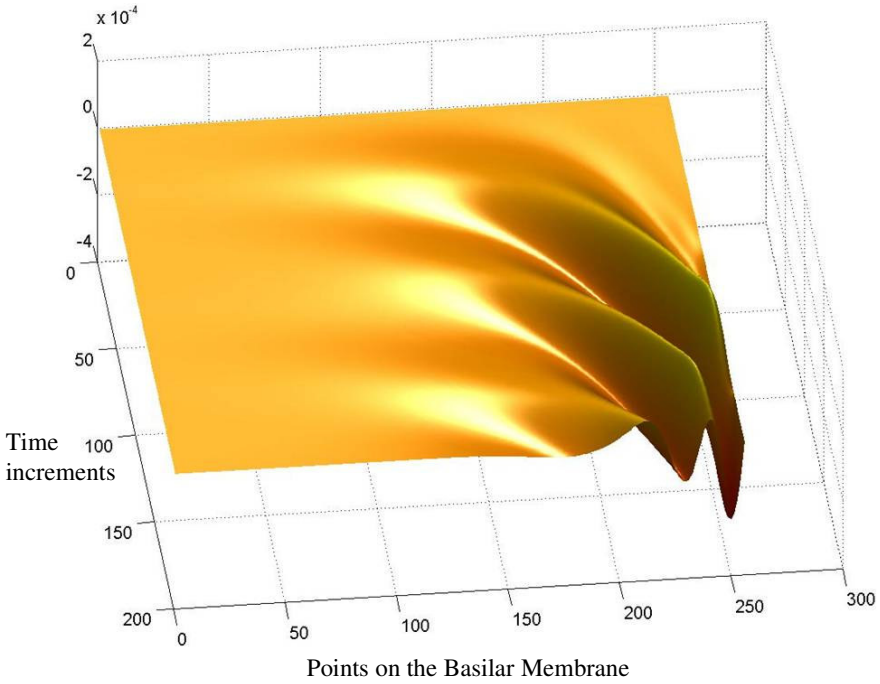

Figure 3.1 Basilar Membrane response to a low-frequency pure tone input

The travelling wave motion can also be seen in Figure 3.2, with the blue, green and red curves indicating the displacement of the basilar membrane on three neighbouring points along the basilar membrane. The red curve represents the displacement of the most apical point. The figure not only shows how the amplitude of the travelling wave increases as it approaches the point of maximal deflection, but also how the wave travels towards the apex, i.e. there is a time delay in the peaks of the deflection waveform.

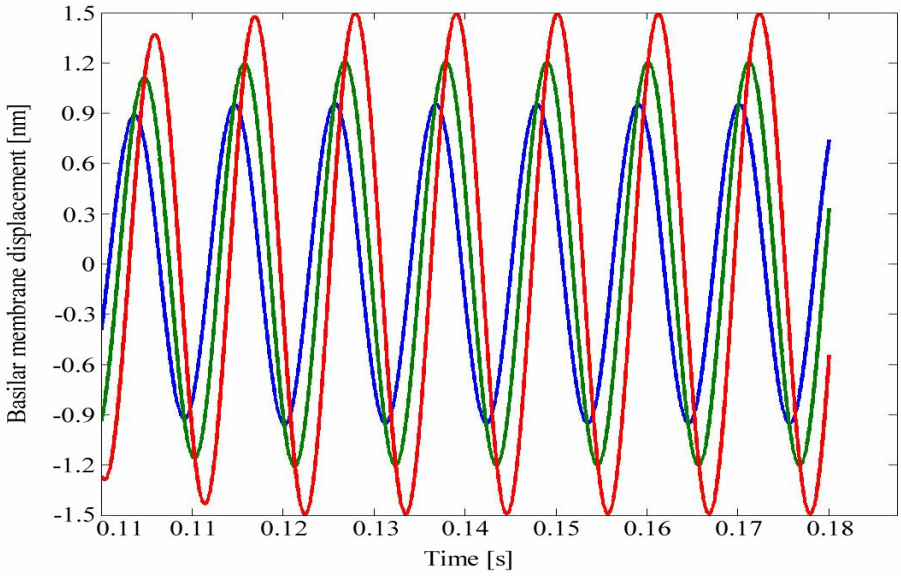

Figure 3.2 Traveling wave on the basilar membrane.

It was therefore possible to implement the model, described by Allen and Sondhi (1979), in Matlab. The output from this model is in the time domain, which allows the output to be implemented as a cochlear implant speech coding strategy by simply translating the basilar membrane displacement to stimulus intensity.

3.5 SUMMARY

This chapter discussed different hydrodynamic models along with their respective strengths and weaknesses. One-, two- and three-dimensional models with solutions in both the time- and frequency-domains has been discussed. A two-dimensional model developed by Allen and Sondhi (1979) was chosen for its time domain solutions which can be implemented as a possible speech coding strategy. Processing incoming sound provides basilar membrane displacement in the time domain. The results obtained were as expected, showing the growth and steep decrease in travelling wave amplitude along the basilar membrane length as well as the increase in phase delay, or deceleration of the

wave, as it approaches the point of maximal deflection. This complex, computational intensive solution, currently limits the implementation of the travelling wave encoding strategy in applications that can process sound prior to its presentation to cochlear implant recipients.

The current restraint on using the travelling wave encoding strategy is its computational intensive solution, currently limiting its implementation to applications that can pre-process the sound before presenting it to the recipient.