MUSIC AND PITCH PERCEPTION OF COCHLEAR IMPLANT RECIPIENTS

Robert Alexander Fearn

B. Elec. Eng (UoW)

A thesis submitted in fulfilment
of the requirements for the degree of
Doctor of Philosophy

School of Physics
Faculty of Science
University of New South Wales

December 2001

To my wife

ABSTRACT

Users of cochlear implants (CIs) often achieve good performance in speech recognition, but report that music sounds unnatural and is often unrecognisable. This thesis reports experiments to determine the relationship between the perception by CI users of parameters related to music (and to tonal languages) and the parameters of the signals that give rise to them.

For electrical stimulation by pulse trains, the perceived pitch decreases approximately linearly with the insertion depth in the cochlear of the place of stimulation. At low rates, perceived pitch also increases approximately logarithmically with the rate of pulse delivery, but it saturates at several hundred pulses per second. The perceived quality of sound is greater when low rates are applied towards the apex of the cochlear, and conversely. The thresholds for temporal jitter discrimination are lower at lower rates ranging from on average 30% at 100 pulses per second (pps) to 50% at 500 pps.

Simulations of the processing of tonal languages were conducted using speech coding strategies currently used for Western languages. These suggest that the resolution provided in the frequency domain is insufficient to allow reliable distinction of words differentiated only by their pitch or pitch contour.

Pitch discrimination was measured in a melodic context using complex and pure stimuli. The difference limens for recognition of change and for recognition of direction of pitch change were similar for pure tones, around 1 semitone at 400Hz, but recognition of direction was much poorer for complex stimuli, around 6 semitones at this frequency. This is explained in terms of the ambiguity of frequency-place mapping of complex signals in current strategies.

In order to improve pitch and music perception, a new coding strategy was implemented that presents rate and place pitch in combination and with the jitter constraints suggested by these studies. Improved pitch discrimination and perceived quality of sound by users are reported.

ACKNOWLEDGEMENTS

I would like to thank my supervisor Associate Professor Joe Wolfe. Joe has always been supportive and encouraging while still being able to let me explore my own ideas. I have appreciated and enjoyed the time I have spent working with Joe and his endless ability to understand and explain things when I cannot. His dedication to research and his students is an inspiration and I am thankful and lucky to have had a supervisor such as this.

This research was funded by an Australian Postgraduate Award (Industry), APA(I), under an agreement between the University of New South Wales and Cochlear Ltd, initiated by Assoc. Prof. Joe Wolfe of UNSW and Dr Paul Carter of Cochlear Ltd. I thank them for instigating this most interesting topic and arranging the funding. Thanks also go to Cochlear Ltd for resources.

I would also like to thank my wonderful wife Theresa, who has provided endless support and encouragement. Thankyou for letting me talk endlessly into the early hours of many a morning, putting up with the despair of failure and being there to enjoy the elation of success.

I would like to thank my family, Mum, Dad, Paula and Katherine for their continued support.

Lastly, I would like to thank all the recipients who tirelessly gave their time to participate in this research. Your desire to hear music has kept me inspired and I hope I can return the fruits of your labour.

Robert Fearn

CERTIFICATE OF ORIGINALITY

I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, nor material which to a substantial extent has been accepted for the award of any other degree or diploma at UNSW or any other educational institute, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at UNSW or elsewhere, is explicitly acknowledged in the thesis.

I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.

Robert A. Fearn

December 2001

ACRONYMS AND ABBREVIATIONS

4IFC	Four Interval Forced Choice
2IFC	Two Interval Forced Choice
ACE	Advanced Combinational Encoder
ADC	Analgue to Digital Converter
AM	Amplitude Modulation
ANSI	American National Standards Institute
BM	Basilar Membrane
BP	BiPolar
BPM	Beats Per Minute
C	Comfort
CD	Compact Disc
CF	Characteristic Frequency
CG	Common Ground
CI	Cochlear Implant
CIR	Cochlear Implant Recipient
CIS	Continuous Interleaved Sampling
CNC	Consonant Nucleus Consonant
CUNY	City University New York
CVC	Consonant Vowel Consonant
dB	Decibel
DCN	Dorsal Cochlear Nucleus
DL	Difference Limen
DLC	Difference Limen for Change
DLF	Difference Limen for Frequency
DLL	Dynamic Linked Library
FB	FilterBank
DSP	Digital Signal Processing
FFT	Fast Fourier Transform
GM	General MIDI
GUI	Graphical User Interface
HB	Happy Birthday
HEI	House Ear Institute

IC	Integrated Circuit
IF5	InterFace 5
IHC	Inner Hair Cell
IIR	Infinite Impulse Response
IPI	Inter Pulse Interval
ISI	Inter Spike Interval
JB	Jingle Bells
JND	Just Noticeable Difference
KHz	Kilohertz
LU	Look Up
MIDI	Musical Instrument Digital Interface
MIPS	Million Instructions Per Second
MP	MonoPolar
MPEAK	Multi Peak
NH	Normal Hearing
OHC	Outer Hair Cell
PC	Personal Computer
PDT	Pitch Discrimination Test
PEST	Parameter Estimation by Sequential Testing
PMMA	Primary Measures of Music Audiation
PPS	Pulses Per Second
PSE	Point of Subjective Equality
RF	Radio Frequency
SAS	Sequential Analogue Stimulation
SD	Standard Deviation
SRMI	Self Rated Musical Interest
SRMS	Self Rated Musical Score
SMSP	Spectral Maxima Sound Processor
SNHL	Sensorineural Hearing Loss
SPEAK	SPectral Peak
T	Threshold
TBM	Three Blind Mice
UD	Up and Down

Contents

List of Figures

Chapter 1
Introduction

> "Following the perception of speech, the appreciation of music is the next commonly expressed requirement by users of cochlear implants" (Stainsby et al., 1997)

1.1 Music

Music is an important part of life. Most people rarely travel through a day without listening to music. This can be by choice, such as playing a CD, listening to the radio, attending a concert, or it can be environmental or incidental music, such as background music at a party, in a restaurant or during a movie.

Why is music an important part of life? What is it about music, whether sung voice or instrument, that can engage our emotions like no other sensory experience? Is it just associative with life experiences or does it go deeper in our makeup? Whatever the answer, it has made music an important part of almost every known culture throughout the world, as entertainment and as art form.

For many cochlear implant recipients, however, music can be a disappointing experience. 83% of recipients report a decline in musical enjoyment and half report that the sound of music is unpleasant or difficult to follow (Tyler et al, 2000). One aim of this doctoral project is to discover why this is the situation. Another aim is more pragmatic. As Shiroma et al. (1997) says "[t]hose cochlear implant users who derive sufficient benefit from the system to recognise speech now have more desire to enjoy music, thus enriching their lives". The second aim of this doctoral project is to help make that possible.

1.2 Scope of thesis

Cochlear implants (CIs) are designed to restore some degree of hearing, to aid in speech perception for those individuals for who have sensorineural hearing loss, for which conventional hearing aids provide little or no benefit. The benefit gained by cochlear implant recipients has been quite remarkable with most being able to hear speech with reasonable perception in many situations and even to conduct ordinary conversations on the telephone. The implants and speech processors that make up the cochlear implant system have focused on the loss of speech and have been developed specifically to maximise the intelligibility of speech. With this success many cochlear implant recipients (CIRs) are now requesting further improvements to other aspects of their lives, such as being able to hear music, making the device smaller and making batteries last longer.

This introduction will provide some background on the way in which the auditory system works, on the various forms of hearing loss, on how a CI works and on their limitations.

The CI system available in Australia is the Nucleus range of devices made by Cochlear Ltd. Most references and descriptions of implants or speech processors will be made to this range of devices, unless otherwise stated.

1.3 Anatomy of auditory system

The auditory system consists of three sections called the outer, middle and inner ear as shown in Figure 1-1. The outer ear consists of the pinna and external auditory meatus or ear canal where sound is directed into the ear canal by the pinna, which also provides some high frequency filtering that aids in directionality (Rice et al., 1995). The ear canal acts as broad filter with a peak at 2700Hz. It also allows the more delicate middle and inner ear to be located inside the head, offering some physical protection. The middle ear consists of the tympanic membrane and ossicular chain. The tympanic membrane, or eardrum, is located at the end of the ear canal and is a thin membrane that vibrates in response to sound waves. Attached to the inner wall of the eardrum is the first member of the ossicles, the malleus (hammer). The malleus is connected to the next member called the incus (anvil), which is connected to the stapes (stirrup). The stapes is attached to the oval window of the cochlea. These bones, a system of mechanical and hydraulic levers, act as impedance matchers for converting air borne sound waves striking the larger tympanic membrane into the vibrations necessary for moving the fluid filled cochlea via the smaller oval window (Green, 1970). There also exists a slow acting, controlled ligament connected to the stapes that can contract to attenuate the transmission preventing very loud sounds entering the cochlea. This reflex mechanism is called the aural reflex. The outer and middle ears provide some filtering with a broad peak of 30dB around 1KHz.

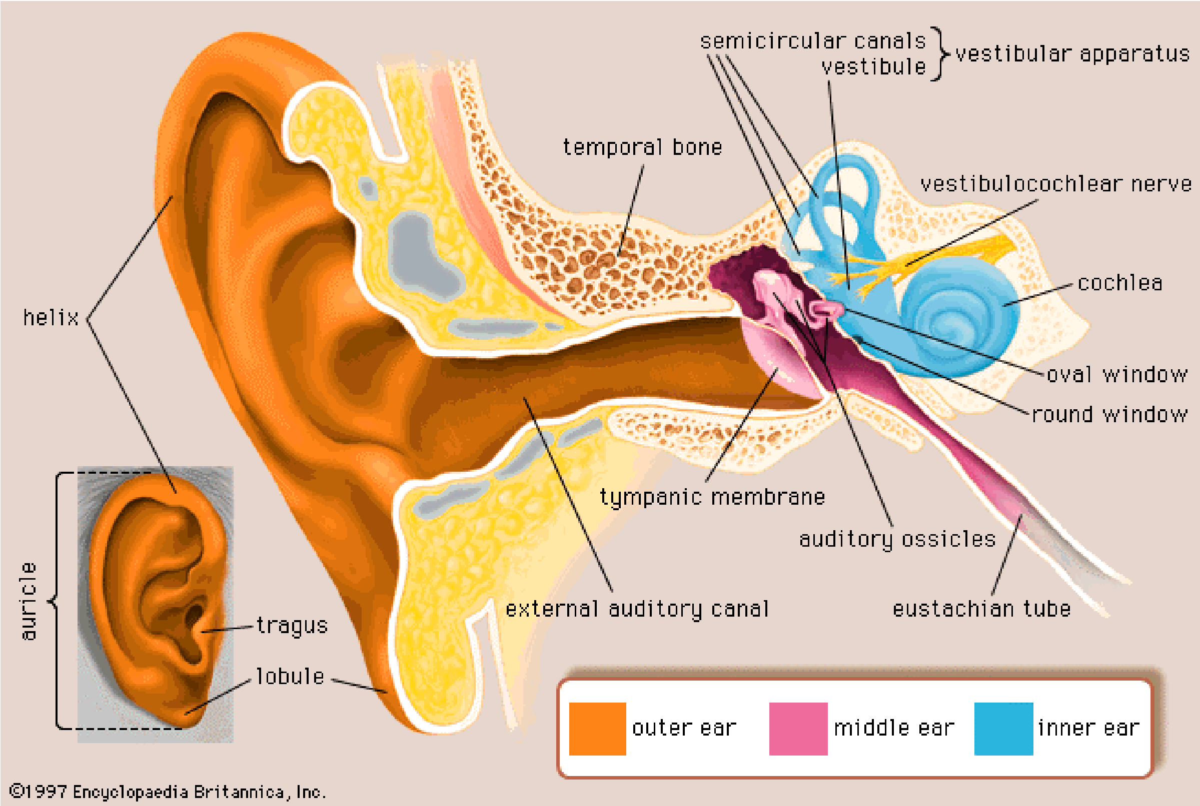


Figure 1-1 The auditory system

The inner ear consists of the cochlea, and the vestibular system responsible for detecting acceleration in space by use of the three semi-circular canals. The cochlea is a coiled, bony, fluid filled structure consisting of three chambers, the scala vestibuli, scala media (cochlear duct), and scala tympani. A cross section is shown in Figure 1-2.

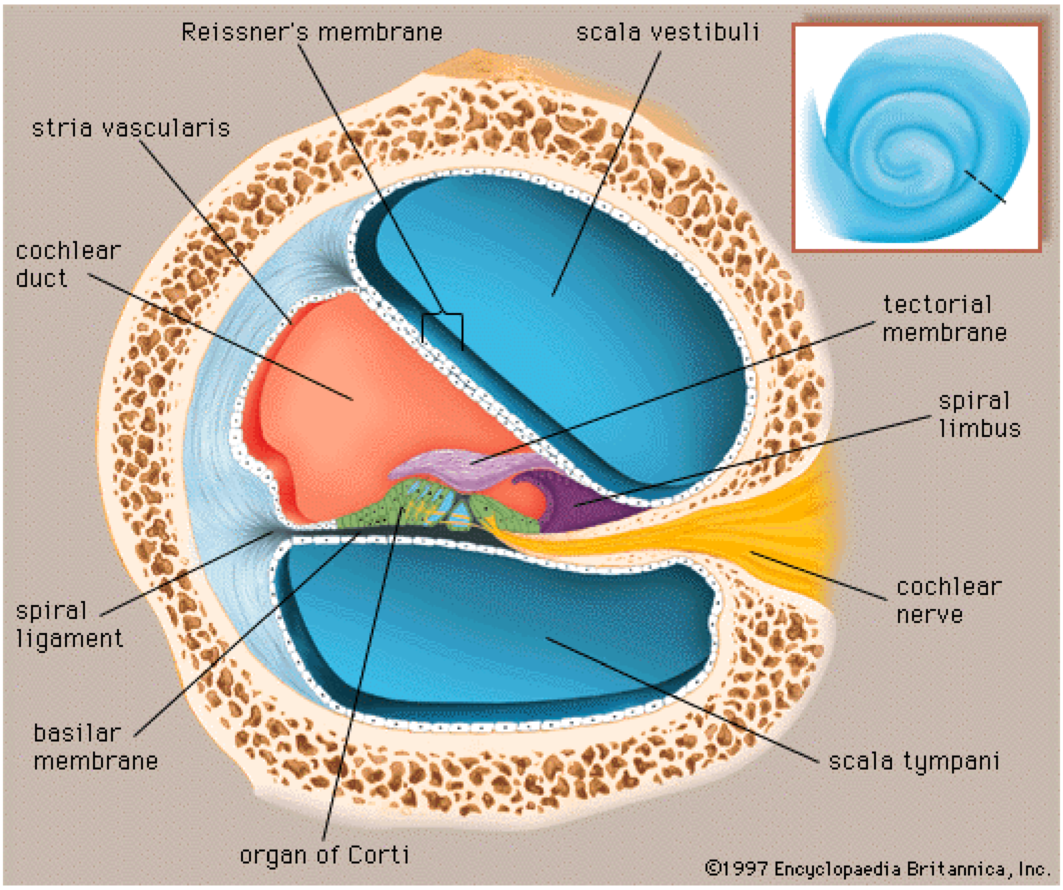


Figure 1-2 Cross Section of the Cochlea

These chambers run from the cochlea's entrance point (base) to the end of the coil (apex). The structure is coiled for almost three complete revolutions for unknown reasons. The scala tympani and vestibuli are connected at the apex of the cochlea by a hole called the heliocetrema and both are filled with a fluid called perilymph, that like most extracellular media, is high in sodium ions and low in potassium ions. The scala media or cochlear duct, lies between the other ducts and is separated from the scala vestibuli by the very thin Reissner's membrane and contains endolymph, a fluid high in potassium ions and low in sodium ions. The endolymph is at a higher electrical potential than the perilymph, which is maintained by ionic pumps in the stria vascularis in outer wall of the scala media. The basilar membrane, which is responsible for converting all the mechanical movements created by the vibration of the oval window by the stapes into neural impulses, is located in the cochlear partition adjacent to the organ of Corti. Along the organ of Corti are hair cells that can be divided into two groups. A single row of inner hair cells (IHC) are located closer to the center of the spiral called the modiolus and are linked to about 95% of the afferent (to the brain) neurons via stereocilia. The outer hair cells (OHC) lie in rows of between three to five of which their stereocilia are arranged in strange yet symmetrical v-shaped patterns as shown Figure 1-3.

Figure 1-3 The single row of inner hair cells and three rows of 'V' shaped outer hair cells

These stereocilia of the OHC are partially embedded in the tectorial membrane and are linked to efferent (from the brain) neurons that can induce motile properties of the outer hair cells by changing their length and raising the tectorial membrane. The OHC motility is extremely fast, able to operate up to 20KHz and usually operates in a rapid fashion (tens of ms), modulating nerve responses to transient acoustic stimulation. They also have a slower mode (tens of seconds), which is able to protect the ear from over stimulation. There are approximately 3000 inner hair cells and 20 000 outer hair cells (Goldstein, 1994). The flexible basilar membrane is 35mm long and varies in width and stiffness, becoming wider and less stiff as it makes its way to the apex. The energy transferred to the fluid maximally vibrates the basilar membrane at locations associated with the frequency. This vibration causes a shear movement between the organ of Corti and the tectorial membrane causing bending of the stereocilia. This in turn opens an ionic channel, allowing sodium ions into the nerve, and causing a nerve action discharge. The spectral analysis is performed by the passive mechanical properties of the basilar membrane and was first observed by von Bekesy who later won the Nobel Prize for Physiology in Medicine in 1961 for research in this field. The high frequencies vibrate regions more basally (close to the window) whereas low frequencies have places of maximal vibration more apically (close to the heliocetrema). This frequency ordering is known as the tonotopic arrangement of the cochlea. The neural pathway begins in the organ of Corti, where nerve fibers flow into the spiral ganglion of the modiolus. They are bundled together to form the eighth nerve which runs into the higher order neural units of the auditory system including the cochlea nucleus inferior colliculus, olivary complex and finally into the primary auditory cortex in the temporal lobe. Compared with other senses, the growth in understanding of the auditory system has been slow, possibly due to the difficult observation of inner working of the cochlea (Nobili, 1998).

The various theories on the coding of pitch and loudness are discussed in the Chapter 2 Literature Review.

1.4 Hearing loss

The frequency range of hearing for normal healthy children is 20 to 20 000 Hz with a dynamic range of approximately 120dB. Frequencies with amplitude variations as low as only a few millionths of a Pascal that displace the eardrum by only the distance equal to the diameter of a hydrogen diameter can be audible. Fletcher-Munson curves show us that frequencies in the region of 1-5KHz have the lowest thresholds (Fletcher and Munson, 1933). Generally, hearing loss is defined as a permanent shift in thresholds of 25dB or more at 500, 1000, 2000 and 4000Hz (Grant, 2000). The many forms of hearing loss can be divided into two categories called conductive and sensorineural. Conductive hearing losses are associated with problems with the outer or middle ear that reduce the transmission of energy to the cochlea. Surgery may correct some of these problems and generally amplification of sound by a hearing aid can be of assistance in conductive hearing losses. Sensorineural hearing losses, SNHL, are usually associated with a reduction in the number of inner hair cells causing a reduction in the ability to transduce the mechanical movement within the cochlea to neural firing. SNHL can also describe higher order neural malfunctions that restrict the neural transmission of sound to the auditory temporal lobe. A CI bypasses the mechanical-neural mechanism of the hair cell by directly stimulating the auditory nerves.

Exposure to noise is considered a major cause of hearing loss. At daily exposure levels of 80dB for 20 years, no damage at 2KHz is found. However at 88dB after 10 years, a drop of 9dB and at 95dB a drop of 15dB is found. Safety requirements suggest that maximum noise levels should not exceed 85dB to 90dB SPL with allowance for very short-term exposure to a maximum of 115 dB. Hearing protection is advised if exposed to above 85dB, especially if long term exposure is expected. Temporary shifts in hearing thresholds are found with exposure to 105dB SPL that recovers in around 1000 minutes. Explosions such as gunfire can damage the ear before the aural reflex can act. These reflexes can act with 40-150 ms and stay active for 1 second. Damage to hearing from noise is cumulative and irreversible (Grant, 2000). Medical conditions that can cause hearing loss include ototoxic drugs, otosclerosis, otitis media, Menier's disease, acoustic neuromas, vascular legions, viral and bacterial infections, head trauma, barotraumas, central nervous system and immune mediated SNHL (Grant, 2000).

It is approximated that 22 million Americans suffer from some form of hearing impairment, which can be categeorised into mild (21-40dB), moderate (41-60dB), moderate-severe (61-70dB), severe (71-89dB) and profound (90dB+). In the USA alone it is estimated that around 500 thousand people have profound hearing loss, around 50% of these are seniors and 40% in the 18 to 64 years age bracket. Five infants per 10 000 are born with severe to profound hearing loss. The hearing impaired are generally financially poorer, have lower educations, and have reduced labor force participation (Blanchfield et al., 1999). The current audiological candidature for a CI for adults is postlingually deafened with severe/profound hearing loss with no speech perception from hearing aids. The ear that had better closed set performance is chosen as it will have more sensitive electrical thresholds, has shorter periods of deprivation, and possibly better radiological characteristics. When there is residual hearing, the poorer ear should be chosen so that hearing aids can be used in conjunction with the CI to aid perception.

1.5 Electrical stimulation of the cochlea

It may have been Volta in the 1800's who first electrically stimulated the cochlea when he discharged about 20 DC volts by means of electrodes inserted into his ears describing the sensation as 'une recousse dans la tate' ("a boom within the head"), followed by a sound similar to that of boiling of thick soup (Volta, 1800). In the 1950's Djourno and Eyries implanted electrodes into the cochlea of a deaf person. The results showed no improvement in speech, but patients could hear sounds and reported it was worthwhile. In 1961, these same researchers fitted a 5 wire electrode into the scala tympani with an external ground wire using a percutaneuous connector, but this was rejected soon after due to redness and swelling before any real trials had begun (House and Urban, 1973). The first successful CIs consisted of a single channel, encoding broadband envelope information by analogue or pulsatile stimulation (House and Berliner, 1982). The first successful multielectrode implants were developed by Clark and Eddington (Clark et al., 1987; Eddington et al., 1978a).

The modern Nucleus implants provide electrical stimulation to the cochlea via an array of 22 electrodes that are spaced 0.75mm apart. These are mounted on a biocompatible silicone tube shown in Figure 1-4. The array is inserted into the cochlea by drilling through the temporal bone into the middle ear, resulting in the geometry sketched in Figure 1-5.

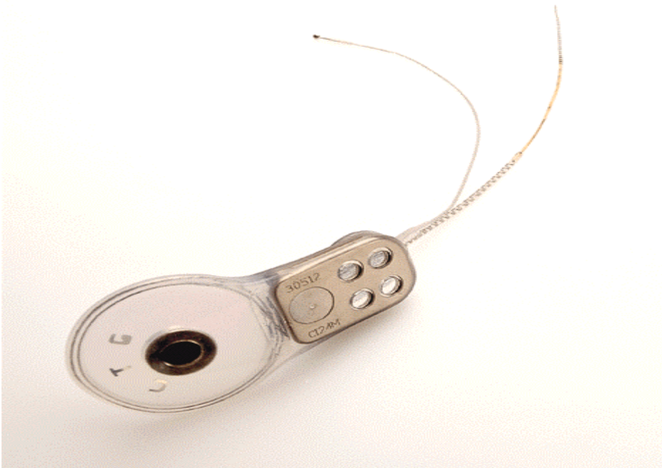


Figure 1-4 Cochlear implant, showing coil, magnet, case housing the integrated circuit, and electrodes on silicone tube. The case and coil are approximately 55mm in length.

Figure 1-5 Cross section of the cochlea with electrode array inserted

A small incision, called a cochleostomy, is made near the round window and the electrode array is fed into the scala tympani, achieving an average insertion depth of around 25mm. The receiver stimulator IC unit that is connected to the array is embedded into the temporal bone. To allow communication with the external equipment a coil containing a magnet is placed under the skin. The external components worn outside the body consist of a communication coil, which is aligned with the internal coil by an opposing magnet, an ear level microphone and a body worn speech processor (or more recently an ear level speech processor). The ear level, directional microphone detects the signal that is then processed by the speech processor and converted into electrical stimulation pulses to be delivered to the selected electrodes.

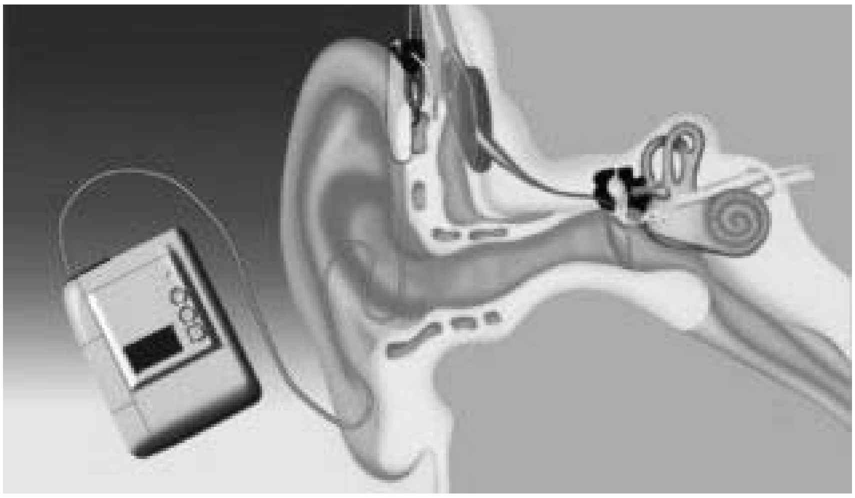


Figure 1-6 Components of the cochlear implant system: The internal coil, integrated circuit and electrode and external microphone and body worn processor.

In the example below, the utterance “choice” is displayed as an acoustic spectrogram and then as an electrodogram. An electrodogram shows the distribution of pulses that are delivered to the electrodes as a function of time and electrode number. Note that the patterns are grossly similar with the high frequency “ch” being represented as stimulation on the high frequency electrodes, the vowel “oi” being represented as stimulation on the low frequency electrodes and finally the “ce” again being represented by high frequency electrodes.

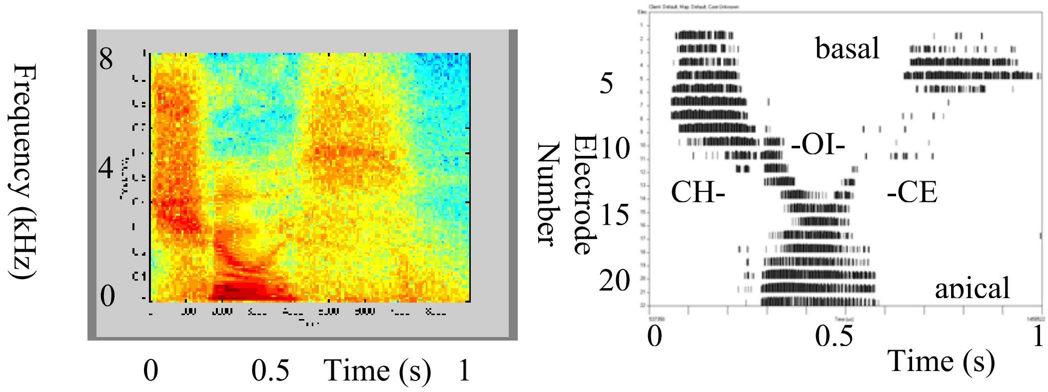


Figure 1-7 Spectrogram and corresponding electrodogram of the utterance "choice"

These current pulses cause depolarisation of the nerve membrane, replacing the excitatory postsynaptic potential from a normal hair cell-nerve. A code is generated based on electrode parameters and this is sent as an amplitude modulated (AM) radio frequency (RF) signal from the external to the internal coil. Power for the internal integrated circuit (IC) and electrodes is derived from the carrier signal. The internal receiver stimulator then decodes the transmitted data and controls the stimulation of each electrode as required. Approximately 3-6 weeks after the operation, the recipient is 'switched on' by an audiologist and the first electrical stimulation is tested. In this programming session, a low amplitude pulse train is applied to an electrode until the threshold of hearing is just reached. This is called the T level. The amplitude is then increased until the maximum comfort level is reached as indicated by the recipient. This is called the C level. These levels are individual for each recipient and the threshold, T, and comfort, C, levels, once gathered during the programming phase, are stored in the external processor so that sound can be mapped between these two levels. The method of converting sound into electrical stimulation pulses is called the speech processing strategy and will be discussed in detail in Chapter 2, the Literature Review.

1.6 Limitations of electrical stimulation

There are major differences between the neural patterns elicited by electrical stimulation compared with normal acoustical stimulation. The CI array has 22 discrete electrodes and thus there is a reduction in selectivity of neural populations with electrical stimulation. Implants have broad spatial extent 1.5-2mm which is an order of magnitude broader than the spatial region over which phase cues occur in the normal cochlea. This may explain the absence of some tonal percepts. The discharges from electrical stimulation tend to be very well phase locked, producing coherent firing in a population of fibers (Bruce, 1997). The neural response to acoustic stimulation is more stochastic with a subgroup of neurons being well modeled by a modulated Poisson process (Goldstein, 1994). Unlike their normal hearing people, CIRs have no background spontaneous firing and there is little independence between adjacent nerve fibers.

Factors affecting performance include etiology for children but not so much for adults, age of onset of deafness, age at implantation, duration of deafness, residual hearing, nerve survival, and device factors. Despite these factors and multitude of variable combinations, most adults achieve above 80% correct in their perception of sentences (NIH Consensus Online, 2000).

1.7 Aims of thesis

Music has three important elements: pitch, timbre, and rhythm. For CIRs, rhythm is transmitted reasonably well to the point where most song recognition is derived from the familiar rhythm. Timbre is perceived not quite as well but adequately enough to recognise major classes of instruments. These two components have analogies to speech that may explain their relatively good perception. Much of the prosody in speech is carried in large temporal modulations while vowel and consonant information is carried in broad spectral peaks and transients. Pitch, however, has less importance in non-tonal languages, where it is used to determine the sex of the speaker and aspects of prosody, such as whether the utterance is a statement or question. It comes as little surprise that, since pitch is not an important part of most non-tonal languages, effective transmission of pitch has been mostly overlooked and consequently pitch discrimination abilities of CIRs are far less than that of normal hearing people.

The aims of this thesis are to investigate the perception of pitch by users of CIs, with a focus in determining ways in which pitch perception can be improved. Pitch perception is not just limited to same/difference tests of pure tones, which has been the subject of most psychophysical research. Pitch perception in a musical context is more complicated. If the results of pitch perception tests are used to predict the ability of CIRs to perceive music, it is desirable to adapt pitch perception tests to a musical format. By using more complex timbres, such as a piano, investigation may reveal more information about music perception than just pure sine waves. Prior research indicates that the type and amount of contextual information makes a significant difference in how accurately normal-hearing subjects perform in pitch testing (Krumhansl, 1991). Consequently, it seems logical that one would also want to consider the effects of contextual cues in pitch testing for CIRs. Furthermore, the recognition of an unfamiliar (new) melody requires that an individual be able to encode a sufficiently accurate perceptual representation of the overall contour of the pitch pattern as well as exact interval size of sequential pitch changes (Gfeller et al., in press). Therefore creating measures that test pitch perception in a melodic context may reveal why CIRs generally have difficulty perceiving previously unfamiliar melodies as unified patterns that can be recognised as 'familiar' (that is, having been heard in prior listening experiences) in subsequent exposures. Often pitch perception has been investigated with regards to the voice pitch information in speech test environments. As most of the languages investigated carry little lexical information in the pitch, it comes as no surprise that speech-processing strategies that aimed to convey voice pitch information saw no improvement over those that do not.

Investigation of tonal languages such as Mandarin or Cantonese, which carry lexical information in the tonal contour, may be more appropriate when investigating pitch in a speech context. The results derived from these studies will be used to determine the optimum parameters for new coding strategies that may improve pitch perception and perhaps improve music perception and appreciation. The practical aim of this thesis project is not just to study pitch perception but also to implement new 'music' coding strategies and to investigate their effect.

1.8 Thesis organisation

This thesis covers several disciplines. The main topic of research is pitch perception but this involves such diverse fields such as psychophysical test design, creation of high level windows programming software for automated self testing facilities and Digital Signal Processing (DSP) in low level assembler language to implement new musical coding strategies. For this reason, the review is divided as follows:

Chapter 2- Literature Review summarises speech coding techniques and the theories regarding the perception of pitch. Also reviewed is how loudness is elicited and perceived and how music has previously been investigated with CIRs. Smaller reviews of jitter sensitivity and perception of tonal languages are provided at the start of chapters 4 and 6 respectively, when those topics are introduced.

Chapter 3 - Temporal Pitch –Place Pitch investigates the effect and interaction of the two main electrical parameters that can elicit pitch, the rate of stimulation and the place of stimulation. Some previous studies have also investigated this field but this chapter aims to provide a comprehensive exploration of this parameter space by varying the rate and place of stimulation through the available and safe limits.

Chapter 4 - Jitter Discrimination is an investigation into the perception of timing constraints of a pulse train specifically observing the thresholds of temporal jitter in pulse trains of various frequencies and locations in the cochlea. The new musical coding schemes described in Chapter 9 require information regarding the perception of jitter at various rates and positions, which this chapter provides.

Chapter 5 – Pitch Pattern Perception aimed to determine pitch discrimination abilities when phrased in a musical context. Melody recognition requires the listener to be able to perceive accurately patterns of connected sound as opposed to isolated stimuli. This task not only requires the discrimination of pitch change but also the direction of pitch change, size of subsequent interval changes, and overall contour of rising and falling pitches. Melody recognition also represents a common real life task relevant to device efficacy in everyday listening circumstances.

Chapter 6 – Tonal Language. This chapter describes an investigation of languages such as Thai and Cantonese, which carry lexical information in the pitch contour of their words. If CIRs have generally poor pitch perception, CIRs whose language is tonal may not only have reduced pitch perception affecting music perception but this may be affecting speech perception. This chapter includes simulations of normal speech coding strategy and an alternative speech coding strategy that aims to provide more voice pitch information, presented to normal hearing listeners to compare the effect of pitch on speech perception.

Chapter 7 -Pitch Discrimination of Pure and Complex Tones: Difference in the Limens. This chapter aims to compare pitch discrimination abilities when the stimulus is pure (sine wave) or complex (e.g. piano). Discrimination will be measured for detection of change in pitch and also detection of the directional change in pitch. The majority of previous investigations have used pure tones only, whereas most real life musical listening experiences involve complex tones of real instruments. Tests involving more complex stimuli may reveal information more relevant for music perception.

Chapter 8 –Measuring Pitch Discrimination Using Alternative Techniques. This chapter aims to determine the differences of pitch discrimination abilities when the pitch is in a musical context. The first test measures the pitch discrimination over a three-octave range. The second test examines the pitch discrimination ability using a standard test devised for normal hearing children.

Chapter 9 - Music-L: A Novel Signal Processing Scheme for Improving Music Perception. A new coding strategy aimed to improve music perception had been proposed by the supervisor of this thesis, colleagues at Cochlear Ltd and the author in the early stages of work on this thesis project. The development of this idea and feasible implementation of two versions of this idea are described in this chapter.

Chapter 10 - Music-L Pilot Trial. This chapter describes some preliminary results obtained with various CIRs using the Music-L coding strategy.

Chapter 11 Conclusions and Suggestions for Further Work. This chapter concludes this thesis project with discussions of the implications of the results and evaluation the strategies.

Chapter 2
Literature Review

This thesis covers a range of topics and a complete literature review at the beginning of this thesis would need to be very broad (and very long). Instead, the background literature immediately relevant to the thesis as a whole will be presented here. More detailed reviews will be given at the start of Chapter 4, Jitter Discrimination and Chapter 6, Tonal Language.

This chapter reviews the speech coding strategies and their development to the present. It also discusses the concept of pitch, the theories that have been proposed to explain pitch perception, and the different electrical parameters that when applied to CIs, affect the pitch percept. The research that has examined the perception of music for users of CIs is also reviewed.

2.1 Speech coding

The strategies that determine the signal processing in the external software or hardware of CIs have been called the ‘speech processing’ strategies. They have been called this because they have been designed to do just that, to process speech. The earliest single channel implant stimulated used a compressed filtered analogue signal. Since then, almost all speech processing strategies that have been used or are currently in use are based on extracting speech features such as formants or splitting the frequency spectrum into sections, whose sizes are chosen to provide maximum speech discrimination. Following is a description of the formant extracting strategies, such as F0F2, F0F1F2, MPEAK, developed by the University of Melbourne and the Nucleus group, the spectral analysis strategies, SMSP and SPEAK, as well as the more recent strategies such as ACE and CIS.

Speech is composed of vibrations created by the vocal chords that are spectrally shaped by resonances in the throat and mouth that can be altered by the tongue. The three main broad spectral peaks created by these resonances are called formants and named F1, F2 and F3. The fundamental frequency of the speech, or voice frequency, is labeled F0.

Simulations of speech coding strategies provide some insight into what a CIR might hear. Simulations are generally created by extracting the amplitude envelopes from the output of a set of bandpass filters (this models the discrete nature of the electrodes). These amplitude envelopes are then used to modulate bandpass filtered noise. The output is then presented to normal hearing listeners and speech perception monitored while parameters of the simulation are modified. Whenever simulations of coding strategies are referred to in this thesis it is implied they were used with normal hearing subjects. These simulations can never fully represent what a CIR does hear as they rarely take into account the reduced dynamic range and neural degradation that CIRs experience.

Throughout this thesis the following notation will be used: pulses per second (pps) will refer to electrical stimulation while Hz will refer to acoustical stimulation.

2.1.1 F0F2

The F0F2 strategy is based upon the extraction of the frequency within the region of the second formant, F2, which ranges from 800 to 2300Hz. The frequency extracted is used to determine which electrode to stimulate. The rate of stimulation is determined by the presence of voicing energy in the region 80 to 400Hz, which is often the fundamental frequency of the voice. If voice information is present, the stimulation rate is at the F0 rate; otherwise the rate fluctuates with signal envelope (Seligman, 1985; Dowell et al., 1985).

2.1.2 F0F1F2

F0F1F2 was developed from research by Y. C. Tong and Peter Blamey, who demonstrated that it was possible to discriminate the sounds created as two stimulating electrodes were changed (Seligman, 1985). This strategy is similar to F0F2 with the addition of the F1 formant. The F1 frequency, 300 to 800Hz, is used to select which apical electrode to stimulate so that there are now two electrodes being stimulated based upon the formants, F1 and F2 (F2 stimulates a basal electrode). The F0F1F2 strategy became commercially available in 1985. Blamey et al. (1985) using simulations of the F2, F0F2 and F0F1F2 strategies with Normally Hearing (NH) subjects found the discourse level tracking using F0F1F2 was superior when used to supplement lip-reading. Significant differences were also found in vowel but not consonant tests using F0F1F2. Researchers have compared F0F2 to F0F1F2 and have found that recognition of monosyllabic words improved by 20% (8% to 28%) and words without context improved by 33% (31% to 64%) using F0F1F2. These researchers suggest the improvements were due to improved transmission of voicing, duration, and envelope features (Tye-Murray et al., 1990).

2.1.3 MultiPeak (MPEAK)

The MPEAK speech processing strategy is an extension of the F0F1F2 strategy with the addition of 3 basal electrodes devoted to high frequency information, electrode 7 (2-2.8 KHz), electrode 4 (2.8-4KHz) and electrode 1 (4KHz+). If voicing was present, the rate of stimulation was at the voice rate F0; otherwise the rate was random around 260Hz. (Whitford et al., 1993; Antognelli et al., 1991). Comparisons of F0F1F2 to MPEAK found improvements due to the MPEAK strategy for vowels 16%, and consonants 17%. They also found that the ‘better’ performing subjects improved from 52% to 80% when tested using open set sentences (Sutter and Schreiner, 1991). Comparing MPEAK to F0F2, it was found MPEAK was better for all 5 subjects tested. 3 subjects also showed significant improvement using F0F1F2 compared with F0F2 (Antognelli et al., 1991).

2.1.4 Spectral Maxima Sound Processor (SMSP) and Spectral Peak (SPEAK)

The SMSP was the first strategy to change direction and move away from formant extraction techniques used previously. It consisted of 16 bandpass filters that were analysed at a rate of 250Hz. The commercial implementation of the SMSP strategy was expanded to use 20 bandpass filters and called SPEAK. This strategy calculated the energy in each of the ‘*m*’ filters and selects the ‘*n*’ largest energy amplitudes, called maxima. It is thus called an ‘*n of m*’, or a peak picking strategy. Once the maxima have been selected, the corresponding electrodes are stimulated at the analysis rate of 250Hz. It was found that some recipients could detect a rate pitch ‘buzzy’ percept due to the 250pps stimulation rate. To reduce this percept the rate of stimulation was randomised around

250pps. Comparing MPEAK with SMSP, participants showed improvement in recognition of vowels, consonants, and all words (McDermott et al., 1992a), specifically, SMSP was better for recognition of closed set vowels (76% to 91%), consonants (59% to 75%), open set monosyllabic words (40% to 57%) and sentences in noise (50% to 79%) (McKay et al., 1992). Results extracted from the graph of Clark (1992a) for recognition of open set monosyllables showed the improvements of speech processing strategies from the 5% achieved by the 1978 prototype to the 15% using the F0F2 in 1985, 32% using F0F1F2 in 1987, 58% using MPEAK in 1990 and 71% using SPEAK in 1994. Clark suggests, “Future improvements in cochlear implant speech processing are most likely to occur with better temporal resolution and place coding of frequency” (Clark, 1992a).

2.1.5 Advanced Combinational Encoders (ACE)

The ACE strategy is an extension of the SPEAK strategy in that it is an *n of m* type strategy. The difference is that ACE generally analyses the spectral energy and stimulates the electrodes at a much higher rate than the SPEAK strategy (250Hz). It has been found that the rate of stimulation and number of maxima selected may be optimised for each subject to obtain the best sound quality and speech scores. The ACE strategy allows many combinations of rate and maxima to be selected, for example, 8 maxima and 900pps/channel or 6 maxima and 720pps/channel. Care must be taken when comparing the ACE strategy due to the many variations that exist among users.

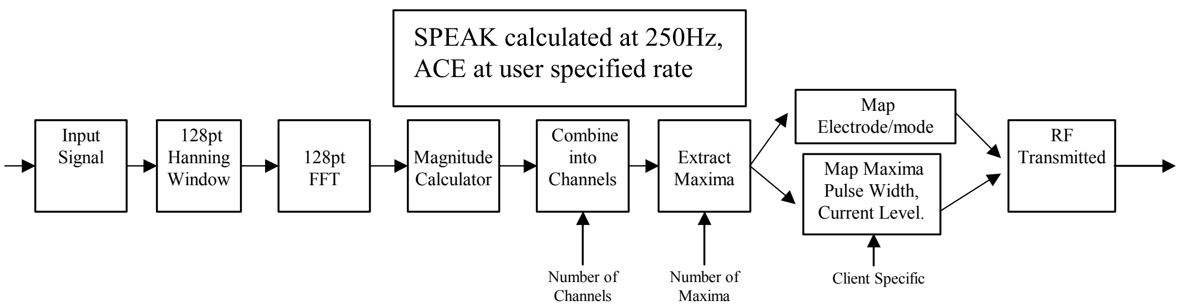


Figure 2-1 Block diagram of ACE and SPEAK signal processing

2.1.6 Continuous Interleaved Sampling (CIS)

The CIS strategy, developed by Lawson et al. (1999), aims to stimulate at the highest rate possible and does not use a maxima selection technique. These parameters result in a higher stimulation rate using fewer electrodes. For example, a CIS strategy may use 12 electrodes (usually equally spaced over the available 22) and stimulate these at a rate of 1200pps/channel. Loizou et al. (1997a) compared SPEAK and CIS. CIS was found to be better for vowel perception in noise when the signal to noise ratio (SNR) was at +5dB. At all levels of SNR tested, the CIS strategy provided better perception of consonants than SPEAK. No differences were found, however, in perception of sentences. Upon further analysis it was found that voicing, place and manner of articulation were higher with CIS than SPEAK. Loizou suggests this was possibly due to higher pulse rate used (800pps) better representing temporal variations in the speech signal. Loizou et al. (1998) compared perception of vowels by men, women, boy and girl groups by CIRs using a 6 channel CIS. They found vowels spoken men easiest to identify, followed by women and boys, with vowels spoken by girls the hardest to perceive.

2.1.7 Simultaneous Analogue stimulation (SAS)

SAS is a strategy that can be produced with the Clarion ABC brand of implant and appears to stimulate each channel using a stimulation signal of alternating current at a predefined rate per channel (Zimmerman-Phillips and Murad, 1999).

2.1.8 Alternate strategies tested

Besides the commercially available speech processing strategies described above there have been many other strategies tested that attempted to improve speech perception by varying electrical parameters. Following is a description of alternate strategies that have been tested and their results. These strategies will be examined to obtain information about how the electrical parameters were varied and how they could be used to improve pitch discrimination.

Tong et al. (1990) used simulations of the F1F2 strategy in comparison with a four-channel filterbank. Vowel perception was found to be similar, however, consonant perception in noise and open set Consonant Nucleus Consonant (CNC) test words was higher for the filterbank strategy than with the formant extraction strategy, F1F2. Jabri and Wang (1996) used neural nets to improve discrimination of sounds using a simulation of a maxima picking type strategy. They reported that it is possible for different sounds to be represented as the same set of chosen maxima. This is due to energy levels that distinguish between being less than the amount required to alter the maxima selected. The neural net inputs were amplitude outputs of a Fast Fourier Transform (FFT) spectral analysis of speech and the output was the channels that were selected. The results of the simulations showed improvement for discrimination of similar sounds using the neural net approach.

Several researchers have experimented with trying to combine temporal information with place pitch information. Lai et al. (1993) tested several hybrid coding strategies based upon the CIS strategy and the F0 parameter in six hybrid combinations. The strategies switched one, two or all of the electrodes between stimulation at voice pitch F0 and a constant rate based on voiced energy or F2 electrode trajectory information. The strategy that switched all electrodes to stimulate at F0 upon detection of voiced energy provided the best male/female identification rate whereas a standard CIS strategy produced chance identification of male and female voices. This study suggests that the temporal presentation of voice pitch was important for detection of pitch related speech features. In his thesis Lai (1990) undertook three studies. In the first study, strategies were tested that chose two electrodes based on F1 and F2 frequencies and stimulated them at rates derived from these frequencies. These frequencies were scaled and/or compressed to convert the F1 and F2 frequencies into lower stimulation rates. Results with two subjects showed that one subject benefited where F1 was compressed to 100-190pps and F2, 190 to 400pps. The other subject benefited when F1 was less compressed (100-250pps) and F2 more compressed (250 to 400pps). In the second study pulse rate interactions were studied by monitoring similarity indexes. Introduction of a background stimulation rate improved some dissimilarities but reduced others for variations of rate and electrode position.

Generally the identification of stimuli with pulse rates below 200pps on the test electrode are more likely to be affected by presence of a background component, whereas above 250pps the background component is less likely to affect identification. These results suggest if rate/place combinations are to be successful then they should be less than 200pps. The third study used multidimensional scaling to measure similarity of sixteen different rate place combinations (4 rates on 4 electrodes). Lai concluded that stimuli with two component pulse rates were perceived to have two perceptual dimensions. The information presented on the more apical electrode pair was more likely to be better received than if presented on a basal electrode. The transmission of information was also better encoded when pulse rates were below 200pps.

To try to improve speech perception, temporal information was modified in speech coding strategies by Jones, reported in his thesis (1994). Comparing MPEAK when the stimulation rate was at a constant rate of 250Hz or at F0 produced no significant differences to speech perception. When the MPEAK strategy was modified by coding the amplitude of F0 using A0 onto the most apical electrode, improvements were found in City University of New York (CUNY) sentences but no other major effects were found. When the SMSP strategy was modified by stimulating the most apical electrode at the F0 rate, while the remaining electrodes were stimulated at a constant rate, no improvements in speech scores were observed. Jones (1994) suggests that perhaps these subjects were exclusively attuned to place based spectral cues and require experience to make use of this rate parameter. Using similar methods, Wallenberg et al. (1990) modified a speech coding strategy by stimulating one channel in a SAS format while the remaining channels were stimulated at a constant rate using pulsatile stimulation. The F2 frequency determined the place of the pulsatile stimulation and 14% improvement was found in speech perception scores compared to single channel stimulation.

In these studies, the common theme was to present more temporal information such as F0 and to test with a non-tonal language such as English. The results showed little or no benefit was gained by adding more pitch related information. This is not surprising as English carries very little information in the pitch of the speech. To identify the benefit of these strategies it may have been more beneficial to use tonal languages that are more dependent on pitch (discussed in Chapter 6) or to measure differences in pitch discrimination ability. The only study that has modified a strategy temporally and observed pitch discrimination was by Peeters et al. (1994), who modified a CIS strategy by modulating the high frequency channels. The results suggested words sounded more natural with improved perception and better discrimination of male and female voices.

Other strategies that have altered the standard frequency analysis techniques have been tested. These vary the analysis rate, window size and stimulation rate dependant on FFT statistics to improve temporal resolution of transient parts of speech and frequency resolution of stationary parts of speech. Results show that some subjects gave better consonant recognition. It was also found that ‘star’ subjects showed better improvement than average subjects (Ouayounet al., 1997).

Irlicht et al. (1995) proposed a scheme that attempts to elicit more natural neural firing patterns in which multiple pulses per period are produced to try to mimic the population period histogram pattern. This however has not been tested with implant recipients. Other strategies in which the frequency analysis is based on wavelets have been proposed but not tested (Jules, 1995).

2.1.9 Parameters influencing speech performance

Several electrical parameters of speech coding strategies have been found to have a consistent effect on speech performance of CIRs. These include the number of channels and dynamic range.

Number of channels

In simulations of speech coding strategies, the number of channels have been found to positively affect speech perception with Faulkner et al. (2000a), suggesting more that 8 channels are required when speech is presented with noise and Dorman et al. (1999), suggesting electrical stimulation of the cochlea is capable of reproducing the high fidelity of the audio acoustic information with six to eight channels. Fishman et al. (1997) found that a single channel gave the poorest discrimination with a dramatic increase noted when increasing from 1 to 4 channels. No difference was found as channels increased between

7, 10 or 20 channels. For CUNY sentence tests, no differences were found between the 4 and 20 channel processor. The researchers suggested this was not due to channel interaction but the test may be limited as the scores approach a ceiling limit.

Dynamic Range

Simulations by Zeng and Galvin (1999a) showed that varying the amplitude compression for speech in quiet and in noise, did not affect phoneme recognition. Reduction of the total dynamic range, however, significantly degraded vowels and phoneme recognition in noise.

Stimulation Rate

Parametric variations of the CIS strategy were studied by Loizou and Poroy (1999) who found pulse rate had a positive effect. 2100pps/channel resulted in improved speech scores compared with less than 800pps/channel. In contrast Vandali et al. (2000), found no difference in speech perception when the pulse rate was varied between 250 and 1615pps.

2.1.10 Other factors influencing speech performance

Blamey et al. (1996b) examined general factors affecting performance of CIRs. They found that duration of deafness had a significant negative effect, increasing after age 60. The age of onset had little effect, the aetiology had a weak effect, but the duration of implantation had a positive effect. It can be implied from this study that the earlier an implant candidate becomes implanted after losing hearing the greater the chance of having higher levels of speech perception. This study also suggests that, the greater experience the CIR has with the implant, the better they will become at using it. Dorman et al. (1994) reports, using analogue stimulation there is a correlation ranging from 0.6 to 0.83 between frequency discrimination ability and speech recognition. He found frequency discrimination was better in the region of F1 rather than F2.

2.2 Pitch, frequency and chroma

> "Pitch is that attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from low to high. Pitch depends mainly on the frequency content of the sound stimulus, but it also depends on the sound pressure and the waveform of the stimulus." (ANSI, 1973)

In a musical sense pitch is used to convey melody and has an upper frequency limit of around 4KHz where the chroma (musical-categorical value, e.g. "c-ness", "d-ness") changes regularly with octave periodicity. For NH listeners, from 4 to 5KHz the chroma is estimated too low with increasing error. Above 5KHz the chroma appears to reach a limit but the tone height continues to rise (Bachem, 1948). The lower limit of musical pitch for pure tones appears to be between 40-50Hz, although the range of pitch perception can be as low as 19Hz (Guttman and Pruzansky, 1962).

2.3 Pitch and pitch coding – a brief history

Cariani (1996) summarises the development of theories on pitch when he states "[p]itch is central to melodic and harmonic structure in music, but despite numerous investigations and spirited debates spanning the last 150 years, the means by which the auditory system represents musical pitch is still poorly understood". The debates are based on different explanations of how pitch is perceived. The three main theories that have been proposed can be grouped into theories based on place, temporal and pattern recognition theories of pitch perception.

2.3.1 Place theory

Helmholtz initially proposed the place theory in 1863 in his book " On the sensation of tone as a physiological basis for the theory of music". Helmoltz suggested that the basilar membrane contained transversely stretched fibers each resonant at different frequencies and thus the cochlea acted as a mechanical spectral analyser (Helmholtz, 1877). This theory was also supported by G.S. Ohm, who had previously formulated 'Ohm's acoustical law' that stated any sound can be decomposed into sine wave components (Plomp, 1964). Helmholtz (1877) also suggested that the eardrum/middle ear were nonlinear and therefore, would produce distortions. The psychoacoustic phenomenon known as the 'missing fundamental' involves the tonal perception of complex waveforms when the listener can perceive a clear sensation of the fundamental frequency is present even if no energy is present at the fundamental frequency. Ohm, in a spirited debated with Seebeck in the 1840's, argued that pitch could only be heard if the stimulus contained energy at the fundamental. Seebeck argued that energy was not required at the fundamental to hear the pitch of this 'missing fundamental' (Wightman, 1981). In 1920 von Bekesy supported Helmholtz's proposed spectrum analyser by observing stimulation of the basilar membrane in cadavers with high intensity signals. He found that the place of maximal excitation on the basilar membrane was related to the frequency applied. Greenwood (1990) proposed a cochlea position to frequency map. This map covering the 35mm of the cochlea proposed 35 critical bands of 1mm each, where spectral components can be heard if they are wider than a critical band. For the human cochlea the formula $Z=(7.24\text{mm})\mathit{ln}(1+f/165\text{Hz})$ gives the place of maximal excitation, Z in mm, from the apex for the frequency, f in Hz (Hartmann, 1996).

2.3.2 Limitations of the place theory

The place theory had problems with explaining the pitch of the missing fundamental. Helmoltz believed this pitch was due to heterodyning caused by distortions in the middle ear/cochlea/basilar membrane. Experiments by Schouten in 1938 and later by Licklider in 1954 showed that the pitch corresponding to the fundamental was not present at the receptor level, suggesting another mechanism must exist for eliciting the pitch percept. Schouten experimented with mixing frequencies that would elicit a missing fundamental, such as 200Hz, with a pure tone at 206Hz. This would be expected to cause beats if the fundamental of 200Hz existed due to distortions. However, no beats were heard. Schouten also experimented with AM showing that although frequencies 1000, 1200 and 1400Hz would elicit a 200Hz pitch, frequencies 1050, 1250, 1450Hz elicit a pitch at 208Hz. If a distortion product were to be present, it should be at 200Hz, not the 208Hz perceived. This shift in pitch from the spacing between harmonics is known as the pitch shift effect (Wightman, 1981). The place theory, which proposes that the perceived pitch is related to the position of maximal excitation of the basilar membrane (BM), also fails to explain how the pitch can remain virtually constant while the sound intensity can alter the location of the maximal stimulation (Zwislocki, 1991; Moller, 1999; Chatterjee and Zwislocki, 1997) by a distance equal to a shift of up to more than an octave (Zwislocki, 1995). This shift has been observed not only in the BM but also in the response of auditory nerve fibers, outer and inner hair cells (Chatterjee and Zwislocki, 1997). It is noted that although the maxima moves basally as sound intensity is increased, the apical region where there is a change from BM vibration to no vibration is unaffected by intensity change (Zwislocki, 1991; Zwislocki, 1995). This suggests a more robust place theory; it may be that the pitch is related to the apical edge of stimulation on the BM. It is also interesting to note that the pitch discrimination difference limen at 2000Hz is 0.2% (Traux, 1999). This pitch difference corresponds to a little more than 10μm between two places of resonance on the basilar membrane. This distance is equal to the distance between two neighboring hair cells, an incredible feat if the place of maximal stimulation is the only mechanism used.

2.3.3 Temporal (fine structure, rate, periodicity) theories

Pythagoras may have been the first to realize a relationship between pitch and frequency noting short strings vibrate more rapidly but it may have been Seebeck experimenting with sirens composed of holed discs who proposed the pitch was determined by the fundamental (Wightman, 1981). In later experiments Seebeck made discs to produce puffs of air with periods of t1 and t2. The fundamental 1/(t1+t2), was heard even though energy at the fundamental was not present. Ohm alleged Seebeck was misled by an illusion and insisted that fundamental energy must be present (Wightman, 1981). Ohm went on to formulate the definition of tone, stating that a tone of frequency f is heard only when the complex sound contains $2a\sin(2\pi(ft+\theta))$ as a component (Plomp, 1964). Wundt suggested that tones give rise to synchronous nerve impulses whose rates determine pitch (Langner, 1992). De Boer and Schouten in the 1940's suggested a 'fine structure' theory involving time intervals between peaks of the acoustical wave. Schouten's 'residue pitch' involves a crude spectral analysis by the BM and the pitch thought to be formulated by neural extraction, especially from the positive part of the peaks in the waveform. The fine structure theory and residue pitch could both explain the missing fundamental but could not explain observed missing fundamental pitch shifts or how a major change in wave shape, such as changing the phase of components, could elicit the same pitch.

The rate of neural firing is a direct relation to the rate of stimulation for rates up to around 1000Hz. This limit is based upon the refractory limit, of around 1ms, of the auditory nerve cells. Above 1000Hz, a population of nerve fibers can follow the rate of stimulation up to 4-5KHz through 'volleying'. Moller (1999) argues temporal coding is supported by the fact that damage to the BM impairing the spectral analysis does not affect speech, but damage to the auditory nerve, impairing the temporal coding, does affect speech perception.

Current temporal models observe Inter-Spike Intervals, ISI, from nerves from all frequency regions. All models look at a predominant interval distribution which, when tested using simulations, can predict a wide range of pitch phenomena. When observing ISI, the most common interval relates to the pitch heard and the ratio of the most common interval to other intervals found, relates to its salience (Cariani, 1996).

One of the problems with the temporal theory is that natural jitter in nerves can cause variations in the order of milliseconds (Pollack, 1968b), which would be thought to impair the discrimination of small differences in frequency that are observed in experiments.

For more details of rate and temporal models, Langer (1992) provides an excellent review.

2.3.4 Pattern recognition models

The third theory of pitch perception is based upon pattern recognition and places less emphasis on the exact mechanism that extracts the frequencies in each signal, but more on the patterns of the frequencies extracted from the input signal. These patterns can then be used to aid in resolving pitches when exposed to similar patterns. Even though these pattern recognition models can explain a wide range of pitch phenomena, no physiological mechanism has yet been identified (Langner, 1992).

In 1956, de Boer proposed a pattern recognition model using a template matching system to find the best harmonic pattern. Goldstein proposed the optimum processor model for central formation of pitch of complex tones (Goldstein, 1973); Terhardt, the learning model (Terhardt, 1974) and Wightman, the pattern transformation model (Wightman, 1981).

In experiments, NH subjects were able to recognize simple musical intervals when two randomly selected harmonics were presented to either the same ear or different ears (one harmonic in each) suggesting that pitch is mediated by a central processor and that non-linearities in the middle and inner ear are non-essential (Houtsma and Goldstein, 1971).

Cohen et al. (1995a) proposed a model of pitch perception using a harmonic weighted sieve, where the pitch is based upon the weighted sum of narrow regions around harmonics of nominated pitch values.

More evidence for pattern recognition being learnt as opposed to inherent, is found from studies with infants who require low frequency energy to hear the pitch of the missing fundamental (Clarkson and Rogers, 1995); unlike adults who require no energy in the lower frequencies to be able to determine the missing fundamental. This suggests that this phenomenon may be learnt and not due to distortions, or fine structures.

Deutsch (1992) has found that the difference of the direction of the pitch of harmonic weighted tritones has been observed between groups having British and Californian residence. She called this phenomenon ‘the tritone paradox’. Shepard (1964) has also created scales of harmonic weighted tones that appear to ascend endlessly. Both of these examples, especially Deutsch’s, support the pattern recognition model of pitch, although Repp has questioned this connection in relation to the spectral envelope used (Repp, 1994).

2.3.5 Combination theories

Licklider (1954) suggests that theories of pitch perception, particularly whether pitch depends upon frequency of the discharges or upon locus of active neural tissue, appear to be both partly correct and the most likely extraction mechanism is a combination of both. His duplex theory consists of a spatial frequency analysis followed by a temporal correlation involving time delay paths feeding into coincidence neurons (Langner, 1992). He also warns that it is clearly futile to try to prove one theory to disprove the other (Licklider, 1954).

Loeb et al. (1983) proposed a spatial cross correlation as a mechanism for pitch perception. They suggest there may be three different mechanisms for pitch extraction operating over different frequency ranges. The first, using place pitch in the 5K-20KHz range, the second, using periodicity rate pitch in the 500 –5KHz range, tracking the stimulation frequency by utilising volleying information while the third mechanism uses pure rate pitch, ranging from 20 to 300-500Hz where nerves are capable of firing on every cycle.

2.3.6 Electrical coding of pitch using time and place with a CIR

Electrical stimulation of the cochlea allows separation of rate and place, unlike acoustical stimulation of the NH ear where they are intrinsically related. This direct control provides an opportunity not only to investigate the optimal parameter space for the CIR but also to improve understanding of how the ear works and to provide evidence relevant to proposed pitch theories. The caveat for this research is that CIRs do not have a normally functioning ear, but one which, has been implanted due to sensoneural hearing loss. This most likely results in reduced nerve survival with sclerosised or ossified cochlea's causing unpredictable current paths. Most adult CIRs have been implanted after period of profound deafness and their memory of pitch may be affected. The results should always be viewed in this light and care must be taken in using these results to explain how the normal ear works.

Two important parameters of electrical stimulation are the ‘rate’ of stimulation and ‘place’ of stimulation. Discrete electrodes determine the place of stimulation, whereas the rate of stimulation is only limited by the capabilities of the processor and implant. Altering the place of stimulation appears to elicit pitch in agreement with the tonotopic arrangement of the cochlea. That is, as the place of stimulation shifts basally from the apical region, the perceived pitch increases. Dowel et al. (1985) also found subjects could rank the perception in basal to apical direction as sharp to dull. Pitch can also be elicited by altering the rate of stimulation, with an increase in rate up to some limit generally producing an increase in pitch (Eddington et al., 1978).

2.3.7 Perceptual dimensions

Simmons et al. (1984) suggested that “using a 100Hz stimulation, there ought to be two pitches. One related to periodicity and one related to place code. Thus, 100Hz should sound like a door buzzer or low hum when apical fibers are stimulated but should sound like a bee buzz in high frequency regions”. In experiments measuring similarity between perceived pitches while varying the rate and place of stimulation researchers have suggested the perceived pitches are perceived separately related to the rate and place dimensions (Tong et al., 1983a; Tong et al., 1983b). The CIRs tested were able to detect when the electrode was rapidly changed but could not detect as well short rapid changes in stimulation rate. Therefore, they suggest the electrode position is better for encoding rapidly changing segmental speech information and pulse rate was better for encoding slowly varying suprasegemental information like that contained in the fundamental frequency (Tong et al., 1983a), although Klein et al. (2001) suggest fundamental frequency is suprasegmental, this issue is still of some contention.

The perceptual dimension can be measured by the dissimilarity index d’ which measures how unlike stimuli are. Tong and Clark (1985) applied stimuli differing in rate and/or place and calculated d’. They found the pitch altered in two dimensions. They also found the musical training or experience with the implant did not appear to account for the variability in pulse rate identification.

2.3.8 Range of rate pitch

Many researchers have investigated the upper limit of rate pitch using electrical stimulation, where an increase in rate will no longer produce an increase in pitch. The upper limit is generally agreed to be less than 1000pps but it is interesting to note the wide range of rate limits reported. This variation may be a result of researchers not reporting how they determine when the upper limit of rate pitch has been found, different experimental techniques, variability of subjects, different types of implants and different stimulation used. Table 2.1 shown below summarises rate limits found.

Rate Limit (pps)	Year	Researchers
400	1978	Eddington et al.
800	1978	Eddington et al.
300	1979	Rosen (cited in Blamey et al.)
300	1984	Simmons et al.
300	1985	Dowell et al.
300	1987	Clark
200-1000	1987	Townshend et al.
200	1994	Wilson et al.
200-600	1985	Tong and Clark
200-1000	1997	Langner

Table 2.1 Limits of rate pitch perception found by various researchers

The lower limit of pitch due to electrical stimulation appears to be around 70Hz, Eddington et al. (1978a) reports that stimulation lower than this rate is perceived as beginning to ‘flutter’.

The upper limit of the rate pitch is most likely related to the refractory period of the auditory nerve, similar to other nerves in the body, being around 1ms (Clark et al., 1990). The neural response to electrical and acoustical stimulation is very different. Phase locking in acoustic hearing is present up to 4KHz; electrical is more synchronous, reaching synchrony 1-2dB above threshold, firing on every cycle up to around 800Hz (cited in McKay et al. (1995a). Acoustical stimulation, however, produces a more stochastic firing pattern with a much greater independence of adjacent fibers. This may limit the ability of electrical stimulation to engage the volleying principle to follow stimulation above 1000pps.

2.3.9 Acoustic - electric matching

Some CIRs have some residual hearing in the non-implanted ear. In pitch matching tasks with these subjects Eddington et al. (1978) reports matches on electrode position of 22mm and a rate of 135pps with an acoustic signal of 125Hz. A 4 mm basal shift while stimulating at the same rate, was matched with an acoustic signal of 195Hz, a 70Hz increase.

2.3.10 Difference limens

Difference limens refer to the smallest difference in pitch that can be discriminated at a specified frequency. Difference limens have been reported by Clark (1992b) to be 6-11% at 90pps, 17-20% at 300pps and greater than 33% above 300pps. Sounds of similar pitch were reported by as many as three electrodes by varying the rate of stimulation, however, the character of sound or timbre was not the same. Difference limens were found to be ~10% when measured at electrical threshold levels but improved to 1-5% when measured at the louder comfort levels (Pfingst et al., 1994). In experiments with monkeys and sinusoidal electrical stimulation, a frequency difference limen of 7% was found at 100pps (17dB Sensation Level, SL) compared to about 30% at 100,300 and 600pps (7-9 dB SL) (Pfingst and Rush, 1987). Hoesel and Clark (1997) have shown that rate difference limens of pulsatile electrical stimulation rapidly rise above 200-500pps ranging from around 10% to 50% for some subjects at 300pps. For sinusoidal electrical stimulation difference limens were 2, 4, 2 and 7% at 100,200,500 and 900 Hz respectively (Merzenich et al., 1973).

2.3.11 Rate pitch to convey music

Eddington (1978a) may have been the first to play pulse rate tunes on 1 electrode. One subject recognised ‘Mary Had a Little Lamb’, ‘Twinkle Twinkle’ and ‘Yankee Doodle’ but not some other tunes, perhaps because the intervals did not appear correct at that frequency. Clark (1992b) played isorythmic tunes comprising of 16-quarter notes. He found subjects could identify some tunes when played on electrode 3 when the tunes were transposed down. Subjects, however, could not recognise tunes on other electrodes but could indicate direction of pitch noticing the intervals appeared smaller than normal. Clark suggests that this may be because rate pitch is occurring in correct place pitch region to sound like the correct intervals, whereas in different regions, there was a greater incongruence between the rate and place percepts.

Pijl (1994) has studied the recognition of interval and tunes while varying the rate of stimulation, probably the most extensively of all researchers in the field. In closed set melody tasks using pulse rates less than 600pps on one electrode, the recognition rates were higher when the tunes were played on apical electrodes. Subjects reported, "apical stimulation sounded more musical than basal stimulation". As with most experiments with CIRs, there are large variations between subjects but he suggests that this is also true of NH when using AM noise and pitch judgment exercises. Pijl also experimented with a novel technique requiring little musical skill that involved adjusting the rate of a second stimulus to match the first two notes of well known melodies, for example, the first two notes of 'Twinkle Twinkle'. He suggests "matching tasks might be applied more readily than musical interval adjustment tasks, particularly with musically unsophisticated subjects". The results suggest that, when rates were kept below 300pps, interval ratios found were consistent with acoustic hearing. When the place of stimulation also changed, this was consistent with expected pitch derived from the cochlea position to frequency maps (Pijl, 1994). In further experiments by Pijl (1997a) subjects adjusted the pulse rates of a comparison stimulus to match that of a fixed reference when the comparison and reference differed in loudness or electrode position. Results showed that, when stimulus rates were of equal amplitude, ~90% of adjustments were within one semitone or less from the target. However, when reference stimulus rate was lower in amplitude, the comparison stimulus (the louder one) was set to a higher pulse rate to compensate for this difference in amplitude. That is, the lower amplitudes were associated with a higher pitch. This effect increased as pulse rate increased. When pitches were matched whose stimulation rates were different and which were applied on different electrodes, he found that although there are gross changes in perception, at low pulse rates, equal rates produced grossly equal pitch percepts even when stimuli were presented on different electrodes.

Shannon (1983) also found that for a 1000pps stimulus, pitch shifts resulting from amplitude variation could exceed those resulting from large variations in electrode position. When examining whether apical regions were more salient, one subject was more accurate in the apical region, but the results were not conclusive. Although it has been shown that rate pitch and place can be separated in different perceptual dimensions Pijl argues that “other data has suggested that it may be possible to compensate for place pitch differences by changes in pulse rate”.

Pijl’s experiments have shown that pulse rate can convey pitch in similar ratios to normal hearing frequencies with regards to being in tune and concludes that rate has an effect and can be used with changes in place to alter pitch.

2.3.12 Pulse rates in speech coding

When determining which electrical parameters to use in speech coding algorithms using a 4-channel device, Simmons et al. (1984) used the results of pitch scaling experiments to determine which formant to place on which channel. They found that the perceived pitch reached a maximum for rates around 300pps and therefore formant frequencies were translated to rates below 300pps for F1 and F2 and high pitches, 2-4KHz, were mapped to more basal electrodes. This coding scheme, however, provided little benefit and subjects achieved more benefit with a strategy similar to F0F1 (discussed earlier).

Wilson et al. (1991) using a hybrid CIS/peak picking strategy with a 4-channel device configured to use the two low frequency channels in a peak picking fashion, stimulating at F0 and F1 rate. The two higher frequency channels were stimulated in a fashion similar to CIS. This hybrid, containing more temporal information, was reported as sounding ‘natural’, especially for music and one subject, a musician, reported percepts had a greater ‘pitch appropriateness’ and musical clarity than compressed analog or CIS alone. The peak picker stimulated at peaks of the bandpass filtered outputs, while the amplitude of the CIS channels modulated a high rate pulse stream. The peak picker and normal CIS showed similar feature transmission of consonants, however the peak picker appeared better for transmission of temporal features like voicing, duration and envelope cues and transmission of vowels. CIS appeared to be better for nasality and place of articulation (Wilson et al., 1991).

Recent four-channel speech coding simulations with NH subjects by Faulkner et al. (2000a) have shown that a strategy where F0 was used to convey voice pitch or noise when the speech was voiceless, gave better perception of glides. Another simulated strategy, using amplitude modulated bandpass noise at modulation frequencies of 32 or 400Hz, however, showed little difference in identification of consonants and vowels. These authors suggest that the use of a fixed rate carrier in the voice range (such as used in the SPEAK strategy) as carrier of a multiband speech envelopes may not be appropriate due to fixed pitch percepts.

Moller (1999) argues that temporal coding is more important than place information for conveying speech information. He suggests that the place code improved the quality of electrically evoked perceptions but were not essential. He proposes that a possible transition of temporal code to place code between 2 and 5 KHz, and that temporal coding is insufficient for high frequency stimuli.

2.3.13 Amplitude modulation

Amplitude Modulation (AM) of pulsatile stimulation can elicit a pitch percept explained more so by temporal than place pitch theories. The perceived pitch appears to be a function of the carrier and modulation frequencies but may produce unpredictable pitches when the carrier rate is less than 5000pps. When the carrier rate was 5000pps the pitch was determined by the modulation frequency and matched in pitch to the frequency of unmodulated pulse trains reaching an upper limit somewhere between 300 and 600pps (Wilson et al., 1997b; Wilson et al. 1997d). Ritsma and Engel (1964) in experiments with

NH subjects, showed the value of modulation index agreed with the hypothesis of fine structure distance between two peaks of the modulated signal.

McKay and Carlyon (1999) reported that multidimensional scaling tasks with CIRs showed pure rate stimulation varied along one dimension while AM stimulation varied in separate perceptual dimensions according to the modulation and carrier rates with perceptual weighting varied according to modulation depth.

McKay et al. (1994a) found that CIRs could rank modulation frequencies of 150Hz and 200Hz well when the carrier was more than 800pps, otherwise the pitch was affected by the harmonic relationship between the modulation and carrier frequency. These same researchers (McKay et al., 1995a) observed the effect of pitch of modulation depth compared to unmodulated pulse train. The results showed that the matched rate fell exponentially from a value close to carrier rate towards a value equal to the modulation frequency, as modulation depth increased. They proposed a model where the equivalent pitch was predicted by a weighted average of modulation and carried frequencies. When the carrier frequency was greater than about 700Hz, where firing on very pulse is probable, the model became invalid and could not predict the matched pitch.

Zhao and Liang (1995) have found neural units in the Dorsal Cochlea Nuclues (DCN) of guinea pigs that show high phase locking to the modulation frequency in the range 400-1200Hz. They report several types of nerves having high modulation gains that increase exponentially with decreasing modulation depth. They argue that there appears to be mechanisms suited to detecting modulation depths as small as 2-5%.

Perceptual reports by CIRs of AM describe the percept to be ‘smooth and tonal’ when the carrier frequency was 1000pps and the modulation frequency was below 200Hz. However when the modulation frequency was above 300Hz the percept became more complex and above 400Hz appears to separate into two distinct tones (Lawson et al., 1994). Shannon reports that CIRs can detect temporal modulations for frequencies below 300Hz and are most sensitive to modulation frequencies in the 80-100Hz range (Shannon, 1992).

2.3.14 Aggregate stimulation

Different pitch percepts can be elicited by combining stimulation from two or more places where the neural population may overlap (McKay and McDermott, 1996). When separation of the two sites was greater than 3-4mm, subjects no longer perceive the aggregate pulse rate patterns but perceive two separate patterns. As amplitude ratios changed from weighting on one electrode to the other, the pitch moved monotonically between those of each component electrode pair, for up to 3mm separation (McDermott and McKay, 1994a).

2.3.15 Mode of stimulation

Mode of stimulation refers to the arrangement of the active and indifferent electrodes used to deliver a stimulation pulse. In all modes, the stimulation is presented in a fashion to produce minimal DC current by reversing the polarity midway through the stimulation. The most common modes are bipolar, BP, and monopolar, MP. In a BP mode, two electrodes within the array are used to present the stimulation. The spatial extent of BP refers to how far apart the two stimulating electrodes are. For example, neighboring electrodes would be labeled BP, whereas an electrode pair separated by one electrode would be called BP+1 mode and so on. In MP mode, one electrode is within the array and the other electrode is located external to the array. The separate electrode may be located in external tissue or it may be the metal case of the implant. Common Ground, CG, is a mode not used as often, where one electrode is active and all other electrodes are shorted to ground, to provide a return path for the current. The threshold and comfort levels are generally lower and more consistent using MP than BP (Wallenberg et al., 1994). With BP, as spatial separation increases (i.e. BP to BP+1 to BP+2 etc) the threshold decreases approaching that of MP stimulation (Chatterjee, 1999). Other less common modes include tripolar and quadrapolar. Kral et al. (1998) measured potentials of implants in an ionic solution and also in cats and found that tripolar stimulation, with side electrodes stimulating 6dB lower than central electrode, produced the narrowest current spread with the smallest side lobes. BP produced two peaks around each activating electrode while MP produced the broadest potential spread.

In experiments involving pitch estimation, the range of pitch appears to be larger when using BP than when using MP (Busby et al., 1994; Wallenberg et al., 1995, in some cases up to 14% (Wallenberg et al., 1996). These researchers also noted that recipients with more deeply inserted electrodes seemed to have greater ranges of pitch using BP than MP. BP has higher thresholds and is less monotonic than MP, however, BP sometimes can produce more pitch reversals. MP is now the most common mode used due to the lower current required. There are no significant differences for speech scores using the two modes (Wallenberg et al., 1996; Zwolan et al., 1996). One advantage of MP is that lower current levels lead to longer battery life. Another advantage is that T and C levels are more consistent leading to easier programming. This is especially useful for programming for young children. Some CIRs have reported MP sounded "clearer " and "more comfortable" whereas BP sounded sharper and higher pitched (Zwolan et al., 1996).

The author speculates that although MP is more common, BP with its larger pitch range, or tripolar with its narrowest current spread, may be more useful for a strategy that aims to improve music perception.

2.3.16 Pulse width

Pulse width refers to the width of each of the biphasic pulses in the stimulation pattern. Generally, equal loudness can be maintained by keeping the total charge delivered to be constant i.e. as the amplitude is reduced, pulse width must be increased to maintain loudness. McKay and McDermott (1997) report, however that pulse width can produce perceptual changes other than loudness. Clark et al. (1997) report that when comparing pulse widths of 250μsec to 25μsec of equal loudness pulse trains, the 25μsec biphasic pulses appear higher in pitch.

2.3.17 Insertion depth

The insertion depth refers to how far the electrode array is inserted into the cochlea. Faulkner et al. (2000c) used simulations of the CIS strategy with NH subjects to propose that identification of consonants, vowels, and key words decreased as simulated insertion depth decreased. These results suggest that using frequency aligned filters whose most apical electrode is 19mm or less from base of the cochlea may result in some degradation of speech. This result is consistent with real CIR measurements by Bredberg and Lindstrom (1995) who reported that speech perception was poorer with insertion depths of 17mm or less. In contrast, Chen (1999) reported that depth was not correlated to pitch scaling or open set word discrimination at 3 months implantation. Comparing the 22 electrode straight array and a preformed spiral array, Chen reports the straight array displayed a broader range of pitches with less plateaus or reversal than the spiral array, and even though the spiral array did allow a deeper insertion, this did not translate to better pitch range. Whitford et al. (1993) using an adapted MPEAK strategy with changes made to frequency boundaries to produce stimulation of speech in a more natural position of the cochlea based on insertion depth, found changes to frequency boundaries improved sentence scores for all 4 subjects tested.

2.3.18 Pitch and timbre

Changes in pitch can be made by changes in rate or changes in electrode. It is possible these produce changes in timbre as well as musical pitch (McKay and McDermott, 1996).

2.3.19 Computer signal processing techniques to extract pitch

Many attempts have been made to extract a pitch estimate from an acoustical signal using signal processing techniques involving computer programs or dedicated digital signal processors, DSPs. The signal processing techniques used with CIRs, however, have only used basic techniques such as zero crossing detection on low pass filtered speech to extract voice pitch frequencies. It may be possible using more advanced techniques to extract more information to formulate a better representation of pitch than just the speech feature F0. Table 2.2 below is a summary of the techniques that have been used to extract pitch information.

Technique	Year	Researchers
Computer program predicting pitch from frequency component ratios.	1979	Piszczalski and Galler
Extraction of various pitches which may be simultaneously evoked by complex tonal stimuli using an FFT, scanning the maxima and estimating frequencies which, are then examined for masking effects and weighted.	1982	Terhardt et al.
Maximum likelihood of F0 is estimated using an FFT, peak extraction and histogram harmonic matching (with good accuracy).	1991	Doval and Rodet
F0 tracking using pattern recognition method of cross correlation.	1992	Brown
F0 estimation using adaptive filtering and zero crossing techniques.	1994	Ohmura
F0 estimation through harmonic estimation so as to minimise error of trial F0 and measured partials.	1994	Maher and Beauchamp
Wavelet to extract pitch of musical signals.	1999	Fitch and Shabana

Table 2.2 Pitch extraction techniques

These techniques can be used as a basis for designing signal processing strategies that will aim to provide an improved pitch perception. Some of the techniques are simple and require minimal signal processing power such as the zero crossing technique while others are more difficult to implement in real-time, due to memory and processor speed limitations, such as maximum likelihood or wavelet techniques. The chosen signal processing technique to represent pitch is discussed in Chapter 9.

2.4 Loudness

Acoustic

The American National Standards Institute (ANSI) defines loudness as "that intensive attribute of the auditory system in terms of which sounds may be ordered on a scale from soft to loud" (ANSI, 1973).

Steven's law states that loudness is a power function of sound amplitude (Stevens, 1935) but has been found to be a more complicated function of amplitude, frequency and waveshape.

Electric

For analogue electrical stimulation of CIRs, the loudness depends on stimulus frequency, and is an exponential function of stimulus amplitude for frequencies above 300Hz and a power function for frequencies below 300Hz (Zeng and Shannon 1995). For pulsatile stimulation, McKay et al. (1994b) found loudness summation increased orderly with spatial separation between electrode pairs and suggested loudness was due to the distribution of discharge rates along the cochlea. They found loudness increased up to 3mm then reached a plateau. Chatterjee (1999) also measured loudness, *L*, and found it was related to current amplitude *i*, by $L = e^{b.i}$, where b to a first order approximation has a linear relationship based on separation of electrodes. Using electrical stimulation, McKay et al. (1994b) found loudness summation in the cochlea was always positive with the absolute amount of summation varying across subjects and electrode position. Clark et al. (1987) found loudness increased with current level and this was more rapid compared to sound pressure but loudness growth rate due to increase in rate of stimulation rate increase was similar. Using CIRs who had some residual hearing in the non-implanted ear, Zeng and Shannon (1992) found loudness balances between electric and acoustic stimulation. The data follows a linear relation between acoustic decibels and electric microamps, and thus an exponential model was found to predict loudness growth determined by threshold and comfort levels. O'Leary et al. (1995) found the discharge rate of nerve fibers to be dependant on the stimulus current and not the pulse rate until a certain threshold was reached. Above this threshold discharge rates became solely dependant on pulse rate.

2.4.1 Pitch dependence on loudness

In NH subjects it has been found that an increase in loudness below 800Hz is found to decrease the perceived pitch whereas an increase in loudness above 3KHz is found to increase the perceived pitch (Stevens, 1935).

Harris (1952) measured the NH difference limens and found they decreased as loudness increased. He also noted that the lower frequencies have larger difference limens, but are also reduced by an increase in loudness. Harris found a serious difficulty when examining pitch and loudness that is highly relevant to experiments with CIRs. He noted, “[w]hen the two tones are being judged equally loud, they are inevitably being judged for equality in pitch, so that, at the final loudness match, no pitch difference can be detected”.

2.5 Music testing

There have been previous studies that have investigated various aspects of CIR’s perceptions of music. These include investigations into the perception of timbre, melody and rhythm.

Some researchers have developed of music tests where quantitative data can be found in a similar fashion to speech tests.

2.5.1 Music test battery for cochlear implant recipeints

Some previous studies have studied the subjective rating of various musical instruments with the results suggesting stringed instruments were more pleasant than wind instruments. Instrument identification is also much lower (~40% for a closed set exercise of 5 instruments) for CIRs than compared to NH subjects (~95%). The complexity also has more of an effect on CIRs than NH subjects in that CIRs have less enjoyment when there is more than one melody. This suggests the complexity of multiple melodies is not perceived and may actually be detracting from the music. It was also found that tonal tests are much more difficult than rhythm tests in comparison with NH subjects (Schulz and Kerber, 1994).

Shiroma et al. (1997) tested CIRs for using rhythm discrimination, instrument identification, H/L pitch discrimination on piano and identification of melody tests. Results showed that rhythm discrimination was 100% accurate. Instrument identification found that the xylophone was most easily identified; ‘humming’ was identified the least correctly, while discrimination of the timbre of the piano was between that of the xylophone and humming. Pitch discrimination ranged form 22% to 86% (chance 50%), while melody identification ranged from 0% to 60% (chance 10%).

Pijl (1997b) using pure tones with NH subjects and electric pulse trains delivered to CIRs, where subjects had to label intervals flat, in tune or sharp, found for low pulse rates (84pps to 119pps), labeling was similar to NH subjects. However, when using normal speech processing strategy with CIRs the pitch interval information does not appear to be available. He also proposes that temporal processing is sufficient for musical pitch and those encoding strategies that restore speech are not necessarily useful with regards to providing the correct melodic interval sizes.

Comparing speech-processing strategies, Tyler et al. (2000) examined music perception while varying parameters such as number of channels (1-8) and stimulation rates. Subjects were asked to rate ‘music quality’ from 3 genres. Strategies with 8 channels were rated higher than the single channel condition, while the results when the stimulation rate was varied were not clear. No higher rating was found for low stimulation rate strategies, but only 3 subjects associated high stimulation rate with better quality. They suggest there are two problems associated with music perception:

i. Inaccurate perception of pitch patterns (melody) and

ii. Low quality of the sound.

To improve music perception Lawson et al. (1994) suggests a synthesiser with direct electrical input to CIRs that can deliver electrical stimulation that can be adjusted to match the memory of musical sounds. In this way the music never has to enter the acoustic domain.

Gfeller has researched the area of music perception as perceived by CIRs with normal speech processing strategies probably more than all other researchers combined. She has researched their musical backgrounds and listening habits (Gfeller et al., 2000), extensively investigated timbre recognition and appraisal (Gfeller et al., 1998), and measured how CIRs perceive rhythm and pitch patterns (Gfeller et al., 1997; Gfeller and Lansing, 1991). She has tested CIRs using already available music and pitch tests (Gfeller and Lansing, 1992) and has more recently developed her own test battery and training package specifically designed for use with adult CIRs (Gfeller et al., 1999). Gfeller's studies have differed from previous research in that they tend to emphasise perceptual accuracy and appraisal of connected patterns of musical stimuli, many of which represent "real world" sounds, as opposed to computer-generated, isolated acoustical stimuli.

Gfeller and Lansing (1991; 1992) have used the Primary Measures of Music Audiation, (PMMA) test, consisting of a tonal and rhythm test designed by Gordon for kindergarten to third grade children. The results suggest that although there is much variation among subjects, CIR perceive tonal information much poorer than rhythm information. Qualitative ratings suggest that a compressed analogue signal provides a more pleasing sound quality than does the F0F1F2 strategy (Gfeller and Lansing, 1992). When comparing the MPEAK and F0F1F2 speech processing strategies, it was found although subjects using MPEAK appeared to initially achieve higher scores on the PMMA tonal test than subjects using the F0F1F2 strategy the differences were negligible after 9 months of strategy use. When investigating timbral recognition and appraisal, Gfeller et al. (1998) found CIRs using CIS were less accurate than NH subjects in timbre recognition tasks and that CIRs found some instruments significantly less likeable, such as the trumpet and violin, compared with NH subjects. Examining the musical backgrounds and listening habits of CIRs, it was found that generally they enjoy music much less post-implantation than prior to hearing loss (Gfeller at al., 2000). Gfeller and her associates have developed a computerised training package for use with adult CIRs that consists of pitch sequence perception, song recognition, timbre recognition and appraisal of different genres of music (Gfeller et al., 1999). More recently Gfeller and her associates have studied melody recognition. The CIR showed considerable variation in perception of pure and complex tones and correlations between melody recognition and speech recognition scores were found (Gfeller et al., Cochlear Implant International- in press).

There have been several studies into how pitch is perceived, but as pitch is of little benefit for CIRs who speak non-tonal languages, there has been little attempt to improve the current pitch resolution. Previous attempts to provide temporal information to speech processing strategies have almost solely been tested with non-tonal language speech tests and little attention has been paid to music perception. It is believed that there may have been improvements in pitch discrimination or advantages for CIRs who spoke a tonal language, for these temporally improved speech processing strategies, but these were never tested.

There has been research into such abilities such as frequency resolution of CIRs and also qualitative perception of music with CIR. There does, however, exist a gap between these two fields of research. It is hoped this thesis can traverse that gap and combine the purely scientific approach with the subjective perception of pitch and music. This thesis aims to combine the literature reviewed with investigations undertaken in this thesis to derive the parameters necessary for improving pitch and music perception. Once these parameters are found they can be implemented in a sound processing strategy.

Chapter 3
Temporal Pitch and Place Pitch

3.1 Introduction

Experiments were conducted in which CIRs estimated pitch while the rate and place of stimulation was varied. Stimuli were concentrated in the apical regions of the cochlea array (where anecdotal reports suggest more musical sounds are elicited). The experiment was repeated on several different electrode pairs. Although it can be argued that varying rate and place alters the pitch in two separate dimensions, it has been shown that a dominant pitch judgment can be made even though the sounds are qualitatively different. Participants were also asked to describe each stimulus in terms of quality, to provide information about regions of stimulation of rates that elicit specific pitches that are more pleasant for the listener than other regions eliciting the same pitch.

The results show that, although perceived pitch as a function of place agrees with the tonotopic arrangement of the cochlea, at low stimulation rates changes in the rate also influence perceived pitch. These results will be discussed in relations to neural firing rates and pre-hearing loss neural memory patterns. Parts of this chapter have been published previously (Fearn et al., 1999; Fearn and Wolfe, 2000).

3.2 Background

CIRs are able to receive stimulation with independent control over the rate and place of stimulation, unlike NH for whom rate and place are virtually impossible to separate. Experiments with CIRs offer the possibility to investigate questions about the human auditory system including those about the relative importance of rate and place of stimulation for the perception of pitch.

Several experiments have shown how pitch perception can be elicited by electrical stimulation of the cochlea. The two parameters that most strongly elicit pitch are rate of stimulation and place of stimulation. Other electrical parameters that have a lesser influence on pitch include mode of stimulation and the pulse width of the electrical pulses.

Early researchers, such as Helmholtz, proposed that the mechanical properties of the basilar membrane acted like a set of tuned filters with the membrane vibrating maximally at positions in direct correspondence with the incoming frequency (Helmholtz, 1877). This vibration of the basilar membrane was later observed by von Bekesy adding support for Helmholtz's predictions (von Bekesy, 1960). In experiments designed to determine the relative importance of rate and place in NH subjects, stimuli with a long-term white noise spectrum was used. Pollack (1968) argued that for such stimuli, spectral cues were the only important factor in pitch perception. The interpretations of these experiments were debated by various researchers of the time, but without true separation of rate and place the place-rate argument could never be fully resolved.

With the invention of the CI, several researchers have investigated the relative importance of changes in rate or place of electrical stimulation. Most of the experiments showed that, when rate changed simultaneously with changes in stimulation place, the change in place had the larger effect on the pitch (McDermott and McKay, 1994; McDermott and McKay, 1997; Eddington et al., 1978). Changes in location of stimulation generally agree with the tonotopic arrangement of the cochlea (Townshend et al., 1987; Tong et al., 1983).

Recent experiments, however, have shown that rate can convey pitch information on a single electrode. Isorhythmical melodies can be conveyed to CIRs by changing only the rate with musical intervals corresponding to changes in rate (Pijl, 1994). Pijl has demonstrated that the rate of stimulation can convey musical pitch in a melodic context however, in these studies, the role of the electrode position (place) in the perception of pitch remained unclear (Pijl, 1997).

Pijl later showed, however, that place and rate both contributed to the perceived pitch by demonstrating that stimulation at a more apical electrode can be matched in pitch to a lower rate on a more basal electrode in a pitch matching task (Pijl, 1997). The Characteristic Frequency (CF) of a given auditory nerve fiber is the frequency applied to the outer ear, to which the nerve responds with a maximum response. The CF of an auditory nerve has not been found with electrical stimulation and further, auditory nerve fibers respond equally well to nearly all frequencies given adequate intensity (Merzenich et al., 1973) (Walloch et al., 1973), suggesting that CF was due to mechanical distribution of incoming stimuli largely by the basilar membrane. It is reasonable to suggest that the place pitch percept of CIRs is learnt before the recipient lost their hearing.

It is now generally accepted that the perception of pitch for postlingually deafened adults depends on both rate and place of electrical stimulation. The interaction of rate and place in the pitch percept has not been previously investigated in the detail necessary to develop methods of effectively conveying pitch percepts to CIRs.

Other mechanisms that influence the perception of pitch include a stimulation rate modulated by a slower frequency, non-simultaneous aggregation of two rates (defined below), varying pulse width, varying the amplitude of stimulation and the mode of stimulation. The effect on the percept of pitch of AM pulse trains has been studied by McKay et al. (1994a). They varied the carrier frequency, f_c and modulation frequency, f_m. The associated pitch percept was highly dependent on the relationship between f_c and f_m when the carrier frequency was less than 800 Hz and when the modulation frequency was less than 200Hz. A distinct pitch reversal was noted between i. f_c =500Hz, f_m=150Hz and ii. f_c =700Hz, f_m=200Hz. This may be due to 'residue pitch' that would be produced if the f_m and f_c were interpreted as two separate pitches that higher order auditory mechanisms combine into one ' best' pitch. When the depth of modulation was varied the pitch moved from f_c to f_m as modulation depth increased. The authors proposed that a weighted average of f_m and f_c related the perceived pitch to a matched single rate. This corresponded reasonably well with the weighted average corresponding to the number of neurons firing at each rate. (McKay et al., 1995).

Aggregate summation of rate is possible by applying individual rates on two electrodes when they are less than 3-4mm (McDermott and McKay, 1996). The perceived pitch could be predicted by the inter-pulse intervals in the total excited neural population, although it was noted that different spatial stimulation might produce changes in timbre that altered this percept.

Because of the discrete nature of the electrode array it was once thought that, at a given rate, only discrete 'pitches' could be perceived. It was found, however, that intermediate pitches could be obtained by non-simultaneous dual-electrode stimulation (McDermott and McKay, 1994). By varying the current levels on the two electrodes, a continuum could be created that traversed the gap of the two distinct pitches created by single electrode stimulation. This was explained as due to overlapping neural populations

Changing the pulse width can also modify the pitch: a smaller pulse width produces a higher pitch for the same rate of stimulation (Eddington et al., 1978).

For CIRs perceptual loudness depends on rate, place, current level and pulse width. The actual function is difficult to formulise due to subject variability. When two tones were compared which matched in rate but varied in loudness, the louder sound was chosen as lower in pitch 71% of the time. It has been shown that an increase in rate could be compensated, in a pitch matching experiment, by increasing the loudness. In one case a 20% increase in rate was compensated by a 4dB increase in current level to maintain the same pitch (Townshend et al., 1987). Pijl (1997) experimenting with pitch matching, found that a decrease in loudness could be compensated by a decrease in pulse rate.

The apical region of the cochlea in the normal ear receives the low frequency components of sound due to the mechanical filtering by the basilar membrane. In postlingually deafened adults it has been thought that the apical region is more accustomed to micro-temporal (as opposed to macro-temporal rhythm) information that relates to pitch and thus apical electrical stimulation may sound more musical than basal electrical stimulation. This may be due to the greater degree of congruence between rate and place information (Pijl, 1994).

Sounds were judged more musical when pulse rates were between 100-500pps than at lower or higher rates and also when applied to more apical as opposed to basal electrodes (Pijl et al., 1995). The majority of adult CIRs are postlingually deaf and therefore the higher order auditory cortices will have years of receiving low frequency components of sound in the apical region where single neurons are capable of firing on each stimulus. It has been suggested that low frequency coding by fibers normally receiving high frequency information could present conflicting information about pitch and that this might explain why stimulation at basal positions sound less musical (Pijl et al., 1995).

There is a very wide range of initial ability in pitch discrimination in the general population, and practice can improve this somewhat (Harris, 1952). CIRs who have been deaf from birth or became deaf prelingually have more difficulty distinguishing differences in pitch (Eddington et al., 1978). It has been argued that much of the perception arising from electrical stimulation produced by repetition rate and electrode position can be described in two distinct perceptual dimensions and cannot be ranked along a single pitch continuum (Tong et al., 1983; McKay et al., 1995b; Collins et al., 1997). Other experiments suggest that it may be possible to rank them in order of pitch, even though there is a qualitative difference (Eddington, 1978; Pijl, 1997). It is thought to be more difficult to match pitches when different electrodes are used. Pijl (1997) argues that the differences between pitches of rates on different electrodes may be analogous to the difference perceived by a NH listener judging pitches that differ markedly in timbre.

Limits of frequency discrimination for the normal listener, based on rate or place alone, have been difficult to study because of the virtual impossibility of separating spectral information from temporal information (Dobie and Diller, 1985). CIRs have been investigated for the range of pitch elicited by changes in rate by several investigators and have been found to vary widely among subjects and techniques used. It is generally agreed that the maximum limit is less than 1000 pps. Pulse rates became confused for stimulation rates greater than 200-600pps depending on the subject (Tong and Clark, 1985). Rate was only useful up to 300-600pps (McDermott and McKay, 1997). Frequency changes that produce significant apparent pitch changes occurred only up to 400-600pps (Merzenich et al., 1973). Above 800pps the pitch does not seem to change whereas from 70 to 800pps the perceived pitch becomes higher with stimulation rate (Eddington et al., 1978). Pitch associated with rate up to 300 pps is clear, but above this, useful rate ranges from 200pps to 1000pps depend on the subject (McDermott and

McKay, 1995). It is quite clear from the results of the various researchers above that the reported limits of rate pitch are quite varied, but generally no pitch change is perceived for changes in rate above 1000pps. A single auditory neuron is thought to have an upper firing limit of less than 1000Hz. This physiological limit may be critical for relaying rate pitch information higher that 1000Hz. Frequencies higher than 1000Hz for a NH person, may be conveyed through a population of neurons using a process know as 'volleying' (Gelfand, 1998).

The difference limens (DL) of pitch have also been measured and again there are marked differences among subjects. One subject had a constant DL of 50% of the rate up to 175pps but above 200pps the DL increased. For another subject a much smaller DL of 9% up to 300pps after which the DL increased (Townshend et al., 1987). When comparing across electrode position, DL's were generally less than 15% for rates less than 300pps but above this rate DL's rose sharply and varied among electrodes (McDermott and McKay, 1997). In comparison for a normal listener DL's of 0.2% for frequencies between 1 to 2KHz in acoustic frequency is common (Hartmann, 1996).

The pitch percept for CIRs also depends on the mode of electrical stimulation of electrical stimulation. The modes of stimulation alter the current paths and are defined as:

i. Common Ground, where one electrode is active and the remaining electrodes are connected together, to serve as a reference.

ii. Bipolar, where one electrode is as active and the other is a reference.

iii. Monopolar, where one electrode is active and either the most basal electrode is a reference or more commonly the reference is an electrode located external to the cochlea.

In experiments using a pitch scaling technique, monopolar stimulation, was found to reduce the range of pitches (Busby et al., 1994).

There are a small number of subjects with a CI in one ear that still use a hearing aid with the other ear which has some residual hearing, albeit small and with a reduced frequency range. They report that it helps in localisation of speakers and aid in understanding speech in some situations (personal comments from a user).

Some CIRs with residual hearing in the non-implanted ear have participated in pitch matching experiments where electric and acoustic stimulation were matched. The implant is around 25mm long whereas the cochlea is around 35mm long and thus the electrode array is not fully inserted into the cochlea. The frequencies are compressed before being applied to the electrodes covering the reduced 25mm length. The corresponding electrical stimulation might be expected to correspond to a much higher acoustic signal as the subject is thought to have a ‘memory’ of the place map (Greenwood, 1990). In experiments matching acoustic and electric stimulation, an electrode placed 25mm from the round window stimulated at 200pps was matched to 1500Hz acoustic signal applied to the other ear. When the electrode position changed to 19mm it was matched to 2000Hz. These correspond on Greenwood’s map to 513Hz and 1363Hz respectively (Eddington, 1978). It is, however, considered quite difficult to match an acoustic and electric stimulus due to the qualitative differences (Merzenich et al., 1973).

In a more comprehensive experiment, four frequencies were matched with electrical stimulation on four electrodes. The results are shown in Table 3.1 below.

Acoustic (Hz)	**Electric (pps)**			
125	135	195	279	313
200	228	278	337	385
300	347	403	415	-
400	390	453	495	-
Insertion depth (mm).	22	18	14	10
Equiv. Nucleus electrode	18	13	7	2
Greenwood's Frequency.	**860**	**1600**	**2877**	**5098**

Table 3.1 Results derived from Dorman et al. (1994) for acoustic and electric pitch matching. The electrode positions have been translated in the Nucleus electrode numbers. The corresponding frequencies for electrode positions derived form Greenwood's map (1990) can be found in the last row of the table.

In these experiments the electrical stimulation was described as having two distinct components. One was a low 'buzzy' frequency, which they used in the matching; the other was much higher and less 'buzzy' and "more nearly a tone" (Dorman et al., 1994). A small number of cases have even reported that acoustic and electric stimuli sound identical (Blamey et al., 1994) but this is not true of the wider CI community.

3.3 Aims

The aims of the experiments described in this chapter are to measure how the rate and place affect the perception of pitch when:

i. Stimulation is concentrated in the apical region.

ii. When stimulation covers the apical to basal region.

This chapter also aims to identify regions and rates of stimulation that can be used to elicit more pleasant musical sound perception.

3.4 Methods

3.4.1 Subjects

Six postlingually deafened adults volunteered as participants for this study. Further participant details can be found in Table 3.2. All participants have been implanted with a CI22M Cochlear Ltd. multichannel device. The implanted device consists of an intracochlear array of 22 active electrodes spaced 0.75mm apart. These electrodes are activated by an implanted receiver-stimulator. A subcutaneous RF link connects the implanted device to the external sound processor. The sound processor the participants use is the SPECTRA-22 ™ or the SPrint™ both programmed with the SPEAK™ coding strategy. The electrodes are numbered from 1 to 22 in a basal to apical direction. The same electrode pairs that the subjects normally use in their speech processing strategies were stimulated in this experiment. Subjects all use the BP+1 mode of stimulation where the electrode pair stimulated is separated by one non-active electrode. This same mode was used for all subjects in this experiment. Subjects participated in two test sessions lasting one-two hours each.

Participant	Sex	Age	Age at Onset	Months Experience	Aetiology	SRMI	SRMS
S1	F	69	20's	132	Unknown	2	2
S2	F	45	30	120	Trauma	2.5	3
S3	F	51	39	144	Ot. Media	3	2
S4	M	35	L0/R30	24	Cogans	2	2.5
S5	F	71	5	38	Unknown	2	2.5
S6	M	72	45	132	Pneumon.	2	3

Table 3.2 Participant Subject Details.

The columns labeled Self Rated Musical Interest (SRMI) and Self Rated Music Score (SRMS) are self-assigned values from a section of a questionnaire below:

Self-Rated Music Score:

1. Music is unpleasant to listen to.
2. Music is OK to listen to but I can't pick the tune or the instruments.
3. Music is pleasant to listen to and I can pick the tune and the instruments.

Self-Rated Musical Interest:

1. I have no interest in music.
2. I have an interest in music.
3. I have played a musical instrument and/or sung solo or in a choir.

3.4.2 Testing software

The author wrote testing software using the Microsoft's Visual Basic programming language. The graphical user interface that the researcher uses is shown in Figure 3-1.

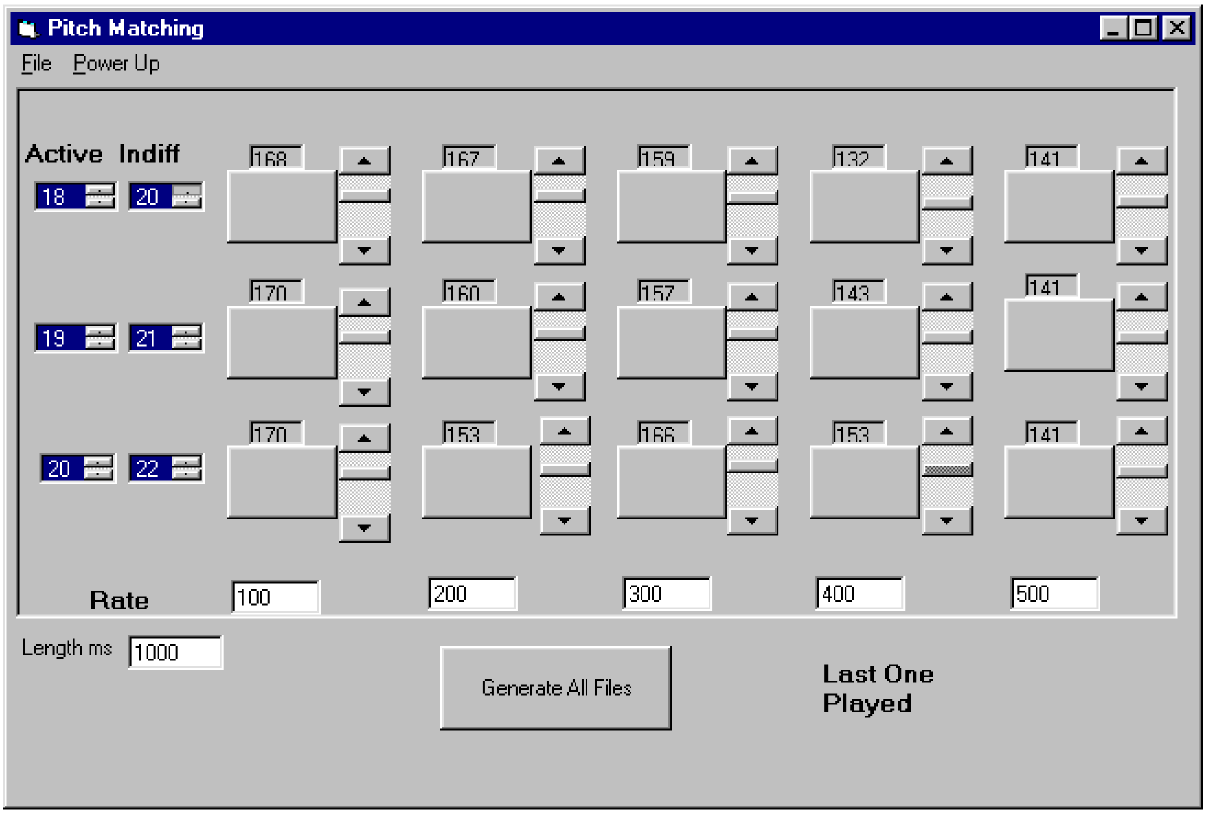


Figure 3-1 Screen-shot of the GUI for the entering of parameters and control of stimulation.

This software allowed the researcher to enter all the parameters required for the experiment, such as electrode, rate and amplitude information. This software communicated with a Dynamic Linked Library (DLL) file written by the author in Borland's C programming language, which received stimulation information and communicated via the computer IF5 card and external PCI unit, with the SPrint™ processor. The Sprint processor had 'stimulation receiving and formatting' software running, also written by the author, to stimulate the desired electrode at the desired rate and amplitude. A block diagram of the system set up is shown in Figure 3-2.

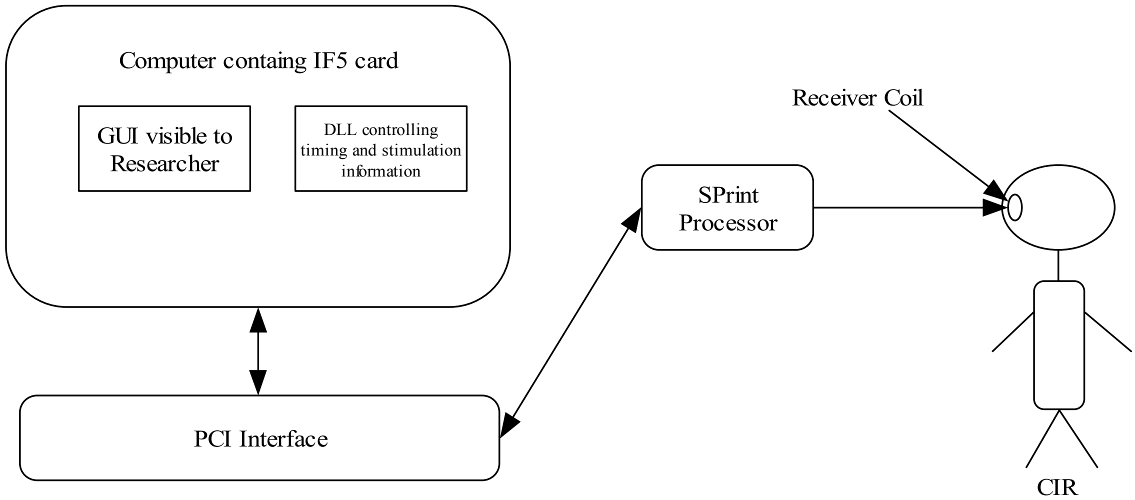


Figure 3-2 Block diagram of system used for this experiment.

3.4.3 Stimuli

The stimuli for the tasks were 1000 ms pulse trains consisting of biphasic rectangular pulses. Pulse duration for all stimuli was 100 μs per phase with 25 μs inter phase gap. Fifteen stimuli consisting of five different rates on three electrodes were loudness balanced to minimise artifactual cues. The subjects were asked to increase the loudness, which controlled the current level, of each stimulus until they reached "medium-loud" (Cochlear Ltd. loudness scale). Each stimulus was then presented following a central stimulus (middle rate and middle position) and the subject was asked to adjust the test stimulus until equal loudness was achieved. This was repeated until the subject was satisfied with loudness equivalence for all fifteen stimuli.

3.4.4 Pitch scaling

The concept of assigning pitches or notes for a non-musically trained individual is almost an impossible task. An alternate scheme, that requires no musical knowledge except for the concept of the direction of pitch, is the pitch scaling technique. The pitch scaling techniques of Collins et al. (1997) and Busby and Clark (1997) were used here. Subjects assigned a value within 0 to 100 following the presentation of each individual stimulus. Zero corresponded to very low pitch while 100 corresponded to a very high pitch. A picture of a piano was shown to the subject and low and high pitches were shown where they appear on the piano if they needed reassurance of the concept of pitch. Presentation order was randomised to ensure no bias effects were produced. All stimuli were presented twice in a training block before data collection. Final scores were calculated by using a modified mean (Collins et al., 1997) in the following way: seven values were recorded for each stimulus, from the seven values collected for each stimulus, the initial value was deleted to allow for an initial adjustment period, the remaining six values had the highest and lowest scores discarded. The modified mean was then calculated from the remaining four values.

3.4.5 Experiment 1-pitch scaling apical electrodes

This first experiment was designed to observe differences in pitch over small changes in place (0.75mm) and small changes in rate (100pps). Frequencies chosen for the first experiment were: 100, 200, 300, 400 and 500pps, using the most apical electrodes (active/indifferent): 22/20, 21/19 and 20/18.

3.4.6 Experiment 2-pitch scaling apical-basal electrodes

Four of the six participants who undertook the first experiment returned to take part in the second experiment. Frequencies chosen for the second experiment were:

100, 200, 400, 800 and 1000 pps.

Using a broader range (apical to basal) electrodes

22/20, 14/12 and 7/5.

The most basal electrodes (e.g. 1, 2 etc) were excluded for two reasons. The first was to avoid any edge effects near the round window of the cochlea where the implant is inserted that may result in a pitch discontinuity, and secondly, so that all subjects were using common electrodes. This was necessary as not all CIRs can use the low numbered electrodes, because of lesser insertion depths or facial nerve interactions.

3.4.7 Experiment 3 - quality assignments

Following the pitch scaling, each stimulus was presented twice, and the subject was prompted for a description of the quality or timbre of the sound. A list of bipolar quality adjectives, shown in Table 3.3 was developed in collaboration with Kate Gfeller from the University of Iowa and Cochlear Ltd. Gfeller selected these particular adjectives based upon existing lists of empirically tested verbal descriptors (von Bismarck, 1974; Pratt and Doak, 1976) and through CIR questionnaires and surveys. This list was given to the subjects where each pair of bipolar adjectives was separated by a line. The subject would indicate a position on the line continuum that indicated how they felt about each sound. This position was converted to a percentage.

```
                 Quality Ratings

      like |----------|----------| dislike
mechanical|----------|----------|natural
   complex|----------|----------|simple
unpleasant|----------|----------|pleasant
     clear|----------|----------|fuzzy
   musical|----------|----------|not musical
     empty|----------|----------|full
     rough|----------|----------|smooth
   natural|----------|----------|unnatural
 colourful|----------|----------|colourless
    hollow|----------|----------|rich
   compact|----------|----------|scattered
      dull|----------|----------|brilliant
```

Overall Rating of the sound

```
       bad|----------|----------| good
```

Table 3.3 Bipolar quality adjectives

3.5 Results

The data recorded from the pitch scaling experiment showed that an increase in perceived pitch was usually produced by an increase in rate, a shift towards more basal electrodes or both. Graphical results for the six subjects are shown in Figure 3-3. The rate is displayed on a log axis to demonstrate the apparent logarithmic nature of the pitch response. The error bars shown are the standard error.

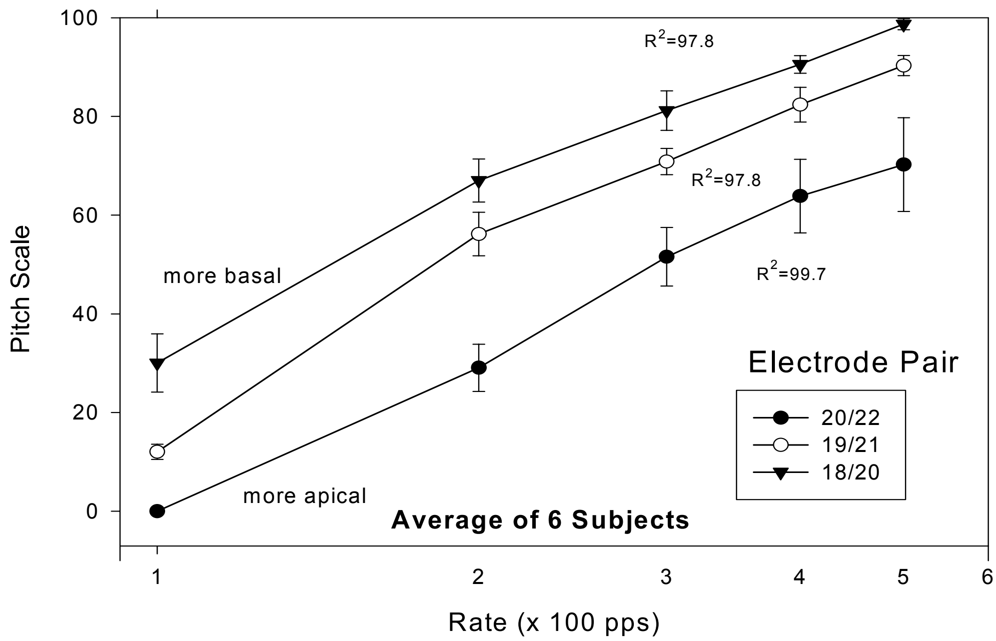


Figure 3-3 Pitch scale as a function of rate of the three most apical electrodes (Error bars shown are calculated as standard error).

A linear regression was performed on the pitch scale value as a function of the log of the frequency on the three plots and the following values were found R^2=99.7 for Electrode20/22, R^2=97.8 for Electrode19/21 and R^2=97.8 for Electrode18/20. These statistics indicated there appears to be strong logarithmic relationship between the numbers chosen in the pitch scaling technique and the rate of stimulation. The lack of musical vocabulary of most subjects and the timbral differences between stimuli make it difficult to relate this pitch scale to real pitch. However, the plots in Figure 3-3 may be formulised as:

$$\text{Pitch Scale} \approx a.(\text{basal displacement}) + b.\log_2(\text{rate})$$

$$\text{Where } a \approx 20\% \text{ /mm}$$

$$b \approx 35\%$$

This may be compared with the logarithmic nature of pitch used in music, e.g. the equal tempered scale. There are some differences among the participants, but there are strong general trends for all six participants. The results are in agreement with the tonotopic arrangement of the cochlea with a basal shift in electrode position producing a higher pitch. There was a contrary example that occurred at one position for a participant who rated electrode 20 higher in pitch than electrode 19 and 18 when the rate was above 260pps. An increase in stimulation rate also consistently produced an increase in pitch. There were no major decreases in pitch as rate increased. There were small, local differences in the perceived pitch at 400pps for participants 4 and 5 on electrodes 19 and 18 respectively. Five out of the six subjects demonstrated that up until 500pps, pitch increased with frequency. Participant 4 perceived large changes in pitch when the rate changed from 100-200pps for all 3 electrodes but above this showed only small changes in perceived pitch as the rate increased. On average the range of pitch perceived by changing the rate of stimulation five-fold on each electrode was a large proportion of the total pitch scale range, being 70.3%, 78.24% and 68.63% of the range for electrodes 20, 19 and 18 respectively. The average change in pitch due to a shift in one electrode position was 15.3% of the total pitch range. This is not surprising since 15% on each of two changes in electrode position leaves ~70% of the used range each of the outer electrodes to convey rate-dependent pitch.

Two of the participants have histories different from those of the remaining four. Participant 5 lost all hearing at age 5 and was implanted at age 67, a period of near silence for 62 years. It might seem surprising that this participant has any pitch scaling ability or even a reliable concept of pitch. This participant also used the smallest range of the pitch scale; around 15%. This may reflect a low confidence in their pitch rating ability. Nevertheless, despite the small range used, the results of this participant show similar trends to the other participants. Participant 4 has an interesting history. The currently implanted ear had never had hearing and the un-implanted ear lost hearing when the subject was 30 years of age. An implant operation on the one healthy cochlea proved unsuccessful due to an acoustic neuroma that damaged the cochlea and associated nerves. Consequently, the electrode array was unable to stimulate residual nerves. The second implant, in the ear that had never heard, was successful. This participant used a large fraction (on average 78.3%) of the pitch scale, when the rate changed from 100pps to 200ppps. However upon increasing the rate from 200pps to 500pps the change in pitch was only, on average, 21.7% of the pitch scale. This suggests that, for this subject, rate pitch does not convey change in pitch much above 200pps. If this result is due to this ear never receiving stimulation until implantation, it may help us to understand how and rate pitch is processed and the extent to which it is a learnt skill.

The second experiment used the same pitch scaling technique but 3 electrodes with a greater spatial separation and stimulation rates up to 1000pps. The frequencies chosen were 100, 200, 400, 800 and 1000pps. The electrodes chosen were 20, 12 and 5 to observe how larger changes in electrode position, than the first experiment, affect the perceived pitch. The results have been normalised for each subject and averaged and are shown in Figure 3-4. All participants' results agreed with the accepted tonotopic arrangement of the cochlea and also reported that increases in rate produced an increase in perceived pitch. Subject 1 reported a large change in pitch with changing rate applied to electrode 20, much less change of pitch with rate at electrode 12 and almost negligible change in pitch with rate at electrode 5. The remaining three subjects did not show such pronounced effects. Subject 3 and 6 reported a general increase in pitch as place moved basally. Subject 4, as in the first experiment, reported changes in pitch with the rate only over the range 100-200pps for all of the electrodes studied. This experiment showed how pitch is affected by large changes in rate and place. These changes are qualitatively similar to those of the first experiment but the effects are larger, presumably because of the larger changes in electrode position. The average range of pitch produced by a 10-fold change in rate at each electrode was 57.4%, 64.2%, and 48.9% for electrode 20, 12, and 5 respectively. The effects of rate changes on pitch are more pronounced in this experiment, presumably because it covered a larger range of rates. Averaging over all electrodes, the changes in pitch produced by a change in rate at the same place was 33% of the pitch scale for the 'rate-octave' from 100-200pps and 9% for the rate-octave 400-800pps. This suggests that rate has a decreasing effect at higher rates. Indeed, from 800-1000pps the change in rate was less than 1%. The average change in pitch place at constant rate was 24% between electrode 20 and electrode 12 and 19% between electrode 12 and 5. Even on a semi log plot (as before) the effect of rate on perceived pitch appears to saturate towards 1KHz, by 500pps they are within 10% of their final value (That is to say that 30% of the change in pitch scaling occurs from over the stimulation octave 100-200pps, while only 10% occurs over the stimulation octave 500-1000pps). On the normalised average data the standard errors were usually less than 5%.

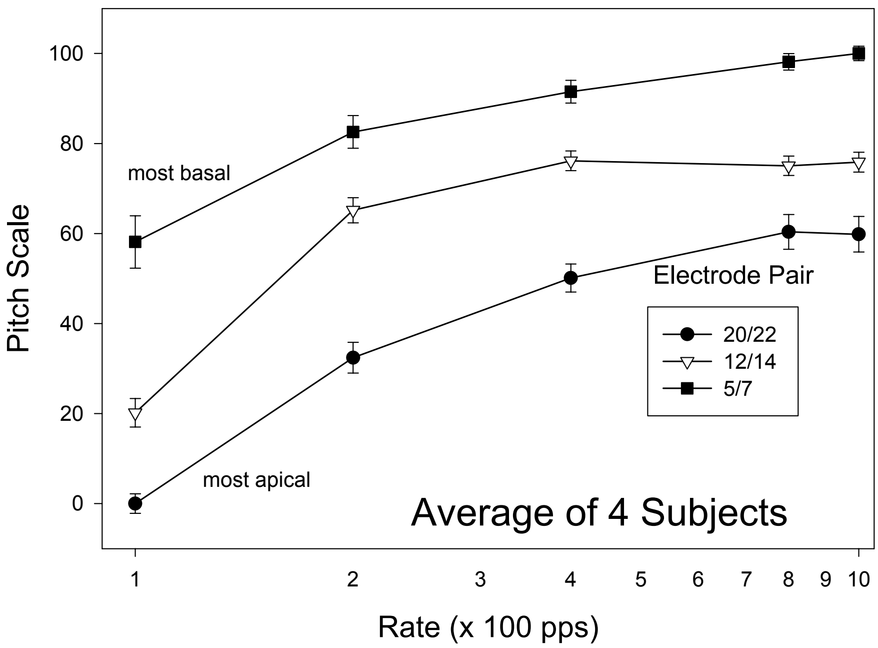


Figure 3-4 Pitch scaling for three spatially separated electrodes for rates 100-1000 pps.

With a much greater range of place investigated, this experiment yields pitch percepts in which rate is proportionally less important. For example, pitch equivalence can be obtained between when electrode 20 is stimulated at 600pps and electrode 5 at 100pps. The graphs shown in Figure 3-3 and Figure 3-4 cannot be directly compared as the range of electrodes and rates are different. The percentage pitch scale also means that absolute comparison is not possible. However, one interesting feature is that the slope below 500pps appears to be different between the two graphs. The slope of the experiment using the smaller range of rates and closer spaced electrodes appears to be almost linear, while the experiment using the larger range of rates and more widely spaced electrodes, is not. This may be due to a number of reasons. Timbral changes associated with electrode position may be greater in the second experiment that may influence pitch choice. The context of a wider range of pitch due to the larger range of rates and electrode positions may also alter the pitch choice.

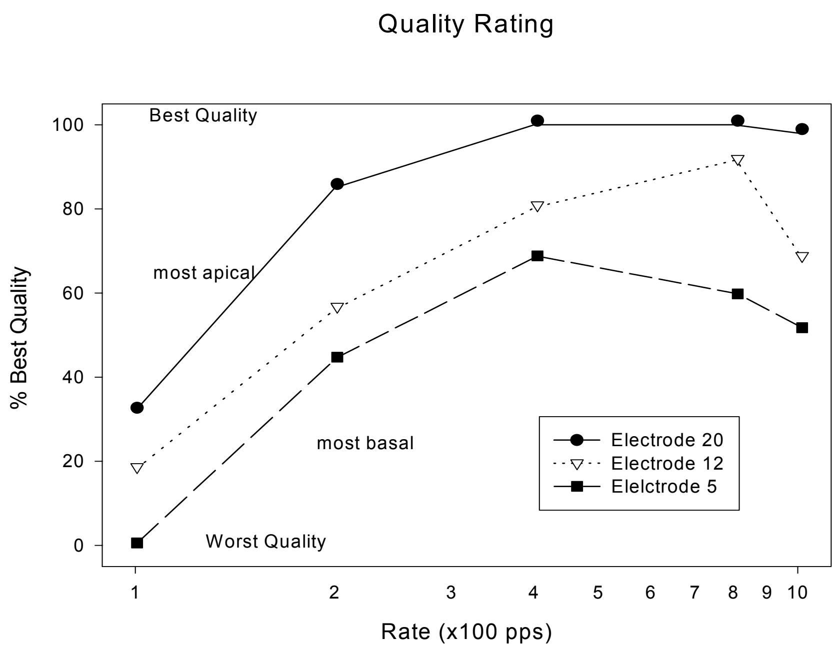


Figure 3-5 Quality rating as a function of rate and place: 100% best quality, 0% poorest quality

All the quality ratings were combined to form a single value ranging from zero (worst) to 100 (best) quality. The results from the first experiment are very similar for each electrode (as expected) and only the results for the second experiment are shown in Figure 3-5. For the first experiment there appears to be an average peak at 'best' quality 300-400pps with 100pps being the worst rated stimuli and 200pps and 500pps equally rated. The second experiment showed stronger trends in the variation across electrodes. For all rates the most apical electrode used, 20, was rated the highest in quality, followed by electrode 12, and then 5, generally decreasing in quality as the position moved basally. Put another way, to achieve any given intermediate level of quality (e.g. 50%) a greater range of rates (greater than 200pps and less than 1000pps) may be used on electrode 20 than may be used on electrode 5 (greater than 300pps and less than 1000pps). The increase in quality due to a shift in the apical direction was significant at the 0.05 level ($p=0.021$) for the quality difference reported between electrode 5 and 12 and also for electrode 12 and 20 ($p=0.039$).

3.6 Discussion

This experiment aimed to determine the relative importance of rate and place in eliciting pitch and to compare the quality of equivalent pitches produced by stimulation of different electrodes. If the pitch depended only on the rate, independent of electrode position the pitch curves of Figure 3-3 and Figure 3-4 would be expected to lie upon one another. If the pitch depended only upon place of stimulation then these figures would be expected to show horizontally separated lines. It is clear that the perceived pitch depends upon a combination of rate and place. At low rates, there is an indication that:

$$a.(\text{basal displacement}) + b.\log_2(\text{rate})$$ may be a suitable approximation

The difference in quality suggests that the apical electrodes produce a sound which is of higher musical quality i.e. more pleasing to the CIR. This is in agreement both with expectation and with anecdotal reports (Pijl, 1997) but is here shown to be statistically significant.

These experiments show that, over a large range of low frequencies, pitch depends strongly on both rate and place of stimulation. They also show that the influence of rate tends to saturate above several hundred Hertz. Thus at low rates, it is possible to elicit the same pitch on different electrodes with different rates even though there may be qualitative differences between the percepts (here expressed as quality). It is envisioned that, once a complete processor coding strategy is developed making use of these results, individual mapping would be undertaken for each subject to find the best rate and place combination to deliver the appropriate quality for the desired pitch. It would also allow the mapping to hold information on equivalent pitches to allow stimuli to be delivered to an alternate site if the first site selected is being used to provide another pitch. Because of the large pitch overlap, these sites can be selected separated by a distance sufficiently large that so no aggregate pitch summation occurs. Thus multiple pitches could be delivered to provide harmonics or even harmony to the CIR, even for small intervals. At rates below several hundred Hertz pitch equivalence can be achieved by altering the rate by 50% to compensate for a change by one electrode.

It seems likely that, for these subjects both the rate and place ability of subjects are due to postlingual auditory memory combined with current speech processor strategies. It is tempting to speculate that, if prelingual deafened recipients participated they might experience less of an effect of rate than that reported by these postlingual recipients. The prelingual subjects' only stimulation of the auditory cortex has been delivered by the speech processing strategy. Currently, these have a constant rate of stimulation and provide most of the information through spectral features.

It may be important to consider the implication for infants with CI who may be losing the ability to utilise rate information if the only stimulus is spectral and at constant rate per channel. If 'rate at place' schemes were used from an early age, it may be possible to achieve better pitch discrimination and make appreciation of music easier.

These results also suggest that, if one could introduce more electrodes in a longer array inserted deeper into the cochlea one might provide an even greater congruence of rate and place information and possibly higher quality and more discriminable sounds.

3.7 Conclusion

The results show at least one of the reasons why a majority of CIRs report music sounds badly with their current processor. The range of pitch available using a constant rate strategy is limited. Strategies currently being developed to convey pitch by changes in position (tonotopic arrangement) as well as stimulation rate (especially 100pps to over 500pps) may have promise for delivering CIRs much more pitch information than the current device provides. These results generally agree with the results from previous researchers such as Tong and Clark (1985), McDermott and McKay (1997), Merzenich et al. (1973 and Eddington et al. (1978). Changes in electrode apically to basally increase the perceive pitch in combination with increasing the perceived pitch by increasing the rate especially up to 500-800pps. These results however, provide a more complete set of data covering a wider range of electrode positions and rates. This has special implications for ‘music’ strategies that may deliver multiple rates and require decisions on rate/electrode combinations to elicit pitch. It would be advantageous to deliver pitch using a combination of rate and place that produce a higher quality or desirable quality of sound where possible. Further work in this area is required to examine the extent to which discriminable pitch can be provided. This may be measured by performing pitch discrimination tasks varying rate and electrode position. Once this information is discovered, it will be left to find whether it is possible to improve pitch discrimination with training using such a variable rate strategy.

These results also support how previous research has described the effect of place of stimulation on pitch. These studies however, provide more detail of the interaction of the rate of stimulation at more places of stimulation, to gain a more thorough understanding of the place-rate-pitch function.

Results from CIRs must always be viewed as a minimal representation of the ability of the normal auditory system. CIRs have severe loss of sensory and neural structures and there are difficulties in participants performing boring, repetitive psychophysical experiments (Dobie and Diller, 1985). Irregular nerve survival patterns and irregular current distributions due to variability in the impedance paths could generate unpredictable neural activity and hence unpredictable pitch percepts (Shannon 1983 cited in Collins et al, 1997). Morphological and physiological characteristics of the auditory pathways have been affected by the mechanisms of the hearing loss, the long term sound deprivation, number of surviving dendrites and ganglion cells, (Busby and Clark, 1997) all lead to variability within subjects. This variability must be taken into account when drawing conclusions involving small numbers of participants and extrapolating to the wider implant community or the general hearing community. Nevertheless, the near linear dependence of pitch on position and its near logarithmic dependence on rate, at low rates, may offer some insight into the mechanism of hearing in general.

Chapter 4
Jitter Discrimination

4.1 Introduction

This study was designed to investigate the discriminability by CIRs of jitter in a pulse train. This information is useful to know for several reasons, some that relate to the implant and the speech processor, and some that relate to neural or perceptual limitations.

Technical constraints on processing.

To design a processor strategy, pulses often must be delayed to avoid coincidence in time so that simultaneous stimulation will not occur. In such cases one must know the maximum time that one pulse in a periodic series can be delayed without being perceived as jitter. Some parts of signals take longer to process than others. One must therefore, know the maximum delay that the signal processing can take before changes in the signal are detectable.

Fundamental questions about perception.

i. Amplitude dependence - is jitter at high amplitudes easier to detect than jitter at low amplitudes?
ii. Rate dependence - is jitter at high pulse rates easier to detect than the same or proportional jitter at low rates?
iii. Place dependence - is jitter at basal locations easier to detect than jitter at apical locations?

This chapter will:

i. Review the experiments done by other researchers with NH subjects and a study where jitter was investigated using two CIRs.
ii. Describe the experiments used in this study with CIRs.
iii. Compare results to those of the reviewed studies and draw implications about which parameters need to be considered when designing signal-processing strategies for CI processors.

4.2 Definition of jitter

Subjectively, jitter is loosely defined as any ‘roughness’ or ‘static’ that can be detected in a signal that would normally be ‘clear’ or ‘smooth’ (Dobie and Dillier, 1985). It has been described as ‘rough’ or resembling a ‘mistuned AM radio’. An objective definition is as follows: variation in the period of a signal, subject to rules defined to produce a given mean frequency and standard deviation of the period. In most research on jitter, the jittered pulse train is created by selecting adjustments to intervals between pulses from a uniform distribution (Pollack, 1968c) or Gaussian distribution (Rosenberg, 1966; Dobie and Dillier, 1985).

4.3 Types of jitter distributions

Several types of jitter distributions have been tested with NH subjects:

i. Uniform distributed jitter (Pollack, 1968a; Pollack, 1968b; Pollack, 1971a; Pollack, 1971b; Dorman et al, 1999) where the interval, ! J is selected randomly over a range defined by the mean Inter-Pulse Interval (IPI) ! some extreme limit J_{max}.
ii. Gaussian distributed jitter (Rosenberg, 1966) where ! J is selected from a guassian distribution centered around a the IPI and the jitter ‘level’ measured in units of Standard Deviations (SD).
iii. Patterns of alternating polarity (Pollack, 1977; Pollack, 1979).
iv. Extreme intervals where the ! J_{max} is randomly added to every interval (Pollack, 1968a; Pollack, 1969) or randomly selected intervals.

v. Jittered pulse trains with filtering such as high pass, low pass and band pass (Pollack, 1971a).

4.4 Pulse trains

In previous studies involving NH subjects the typical pulse trains consists of 10μsec pulses, usually positive going, except for randomly polarity pulse trains, which included both positive and negative pulses. The pulse trains are generated and jitter added. The IPI is kept constant by adding both interval ! J at some stage during the pulse train. The pulse train is then presented ,via headphones, to the subject to be tested. The duration of each stimulus is typically 500ms with an Inter-Stimulus Interval (ISI) of 750ms.

4.5 Psychophysical testing methods

Other researchers have previously used two main psychophysical testing methods. These are Four Interval Forced Choice (4IFC) and Two Interval Forced Choice (2IFC):

i. The most widely used technique (Pollack, 1968a; 1968b; 1971a; 1971b; 1968c; 1969a; 1968d; 1979; 1969b) is the 4IFC which uses the adaptive technique of Parameter Estimation by Sequential Testing (PEST) (Taylor and Creelman, 1967) set to estimate the 50% discrimination level. In the 4IFC techniques the subject is presented with four pulse trains and must nominate which of the four pulse trains is jittered, or more jittered than the other three. The PEST technique is adaptive and modifies the step size between successive presentations based upon the history of responses.

ii. In the 2IFC with an adaptive technique the subject is presented with two intervals of pulse trains and must judge if the stimuli are the same or different, or when the task is discrimination, if the first or second stimulus is higher or lower in pitch. When the distribution of the pulse train was normal, the discriminable levels were usually reported as $\Delta T/T$ as a percentage, where ΔT is the standard deviation of the pulse train with an IPI of T. For other distributions such as Guassian, the level of jitter reported is the (SD or the SD/T.

The discrimination threshold found on the psychometric curve using the two adaptive techniques described above are 50% using the 4IFC and 71% using the 2IFC (Wetherill and Levitt, 1965; Levitt, 1971). The probability of a decrease in level using the 2 down 1 up adaptive staircase method is $P(x)^2$ and the probability of a correct response is 0.5 for the 2AFC and 0.25 for the 4AFC. The point on the psychometric function found at convergence is thus, $\sqrt{0.25}$=50% for the 4AFC and $\sqrt{0.5}$=0.707 $\cong$ 71% as stated above.

4.6 Normal hearing limits of jitter detection

NH limits of jitter discrimination have been found using various distributions mentioned above. The parameters that can be varied within each stimulus are:

i. **n** - the total number of pulses or intervals.
ii. **IPI** - the time (1/frequency) or period.
iii. **J** -the level of jitter.

Pollack (1968a) found limits in J with NH listeners of 0.1% (n=128 IPI=0.3msec), 1.5% (n=32 IPI=3msec) and 15% (n=4 IPI=30msec). In another study, Pollack (1968a) found that the level of discriminable jitter varied over 400% depending on n and IPI. He found pitch discrimination improved with increasing n, especially in the region of 2KHz. When the interval was jittered by a large J, the number of pulses had little effect on the ability to discriminate. In other studies Pollack (1971b; 1968c) explored the relationship of combining amplitude jitter and time jitter and suggested a vector addition model to explain the results when amplitude and time jitter were combined. He found amplitude jitter discrimination levels as low as 0.1% for n=2000 and IPI =250μs. Pollack (1971a) found jitter discrimination levels were worse (higher) for low frequency pulse trains than for high frequency pulse trains. It was suggested that the spectral frequency differences between high frequency spectral components of low frequency pulse trains were too close together to permit effective discrimination. Rosenberg (1966) found that a jitter with absolute magnitude in the order of 1msec had a nearly constant effect on pitch discrimination. Adding low pass noise below 1000Hz to the pulse train tended to make judgments very difficult. He found that the jitter was critical to rendering pulse trains easily discriminable related to a flattening of the spectra at 250Hz. Combining the results from both experiments, he concluded that the most important region for detecting jitter lay between 300 and 1000 Hz, and hypothesised that subjects were utilizing neural volleying patterns to aid discrimination. Pollack (1969a) found that, when broadband noise in the region of 4800-9600Hz was added to jittered pulse trains whose center frequency ranged from 66Hz to 4.3KHz, there was little effect on jitter discrimination levels. However, broadband noise in the region of 20-1200Hz reduced the discrimination ability as the noise level was increased. When noise at 250, 1KHz and 4KHz was added to a 260Hz pulse train, the threshold jitter was 15μsec, 150μsec and 12μsec respectively. This trend also held for frequencies of 66Hz to 4.3KHz. Clearly the noise added at 1KHz dramatically reduced the discrimination ability, suggesting the importance of the 1KHz region for jitter discrimination for this 260Hz rate. Pollack also found that, above 1KHz, increasing the signal amplitude level allowed lower discrimination thresholds whereas below 1KHz, increasing the signal amplitude level above 45dB produced higher discrimination thresholds. The criterion for the transition point in these experiments was when 66% of judgments were 'non-discriminable.' Pollack suggests the 1KHz region has a strong role in jitter detection.

4.7 Cochlear implant recipient limits of jitter detection

In experiments with two single channel CIRs, Dobie and Dillier (1985) used Gaussian distributed jitter. Subjects reported jittered pulse trains as 'rough' and could distinguish the 'rough-smooth' dimension differently from loudness and pitch. Both subjects showed slightly worse discrimination at lower frequencies.

Freq(Hz)	CI Subject 1	CI Subject2	Normal Subject1
80	27.5	-	1.47
125	15.6	19.2	1.04
250	19.2	32.5	0.83
500	13.5	18.3	0.83
750	30.0	-	1.0
1000	5.0	14.2	1.0

Table 4.1 Jitter threshold results taken from Dobie and Dillier (1985) for two Cochlear Implant Recipients and one NH subject. Thresholds expressed as the standard deviation divided by the Inter Pulse Interval, SD/IPI.

4.8 Natural jitter in nerves

In a 2.5KHz pulse train presented to NH subjects, very small changes in the order of 2μs are easily observed. Individual auditory nerve fibers, however, demonstrate firing variations in the order of milliseconds (Pollack, 1968b). Results from jitter experiments suggest that, to imitate the temporal firing pattern of single auditory nerve fibers, independent stochastic pulsatile stimulation would need to be supplied to individual neurons (Dobie and Dillier, 1985). The propagation time across a neural synapse is of the order of 50μs, the first stage of the cortical pathway has a minimum of four synapses, and thus temporal resolution of around 200μs. NH subjects, when attempting to detect jittered pulse trains, are able to achieve lower thresholds when the pulsatile presentations contain large n and small IPI's. In low frequency regions around 100Hz, the NH subject's jitter thresholds are of the order of 1% which is thought to be too precise to be achieved by individual temporal units with inherent imprecision of 30 -150% (Pollack, 1968a).

4.9 Detection of jitter

One theory proposed by Dobie and Dillier (1985) for jitter detection is based on unit arrays such as Licklider's (1951) neural correlator of direct and delayed spike information. Another theory is based on models of oscillators with varying positive feedback lengths. It appears that the receiver of many individual units may be much more informed and therefore make better judgment of jitter than individual units can do. Receivers must exist that can detect small differences in average arrival times because the time differences that can be detected by in binaural arrival times are of the order of μs (Pollack, 1968a).

4.10 Comparison of normal hearing jittered pulse trains and cochlear implant jittered pulse trains

The majority of the literature report tests involving presentation of trains of short pulses to subjects via headphones. The pulse train is rich in harmonics at all multiples of the fundamental frequency (unlike a square wave signal which has only odd harmonics). When presented to the ear, this will produce an excitation pattern over a large region of the basilar membrane. When pulse trains are electrically delivered to CIRs, they are usually delivered to one electrode at a time and thus more limited in spatial extent and so are analysed only by the immediate surrounding nerves. An equivalent signal in the time domain may be a narrow band filtered noise that has each period jittered.

It is assumed that the threshold levels for CIRs may be much higher. This assumption is based on the fact the due to the discrete nature of the CI electrode allow there is a reduction in the total number of independent sites that can be stimulated and that at these sites the spatial spread is larger for the CIRs than NH subjects.

4.11 Experimental work

4.11.1 Aim

The aim of this experiment was to investigate timing “jitter” discrimination function of CIRs as a function of frequency and electrode position. This function includes both the jitter discrimination threshold and difference limen. The jitter discrimination threshold is the point at which the jitter is detected at a rate equal to 50% successful discrimination. The difference limen is the smallest detectable difference between two stimuli that can be detected 50% of the time. The difference limen is calculated as the Probable Error (PE), or .67 times the standard deviation of the values of the variable stimulus evoking judgments of equality. The difference limen is a measure of the slope of the discrimination psychometric curve. A small limen indicates a slow transition from non-detection to detection, whereas a large limen indicates a fast transition from non-detection to detection of the jitter.

The hypotheses to be tested are:

i. Threshold level is dependent on frequency or stimulation rate.
ii. Threshold is not dependent on position or electrode.
iii. Difference limens are dependent on rate, i.e. the slope of the threshold function is dependent on rate.
iv. Difference limens are not dependent on position or electrode.

The first hypothesis is supported by the results on NH subjects in whom the threshold is a function of the stimulus frequency. For NH subjects, the literature suggests that the thresholds are much larger for low frequency pulse trains than for high frequency pulse trains (Pollack, 1971a). One may expect that the absolute levels detected by the CIRs will by larger than those detected by NH subjects but may follow a similar trend in relation to frequency.

The second hypothesis proposes that there should be no difference between the thresholds as the same rate is applied to different electrodes. This is suggested in spite of the literature reports that the 1KHz region on the basilar membrane is very important for jitter discrimination of NH listeners (see section 4.6, this chapter). This proposed difference is suggested by the observation that, as the electrical signal is applied to just one location on the BM of CIR, the position of that electrode is not critical.

The third hypothesis also follows NH reports that if there is a dependence of threshold on frequency it might also be true that the slope of the psychometric threshold function may also be dependant on frequency.

The fourth hypothesis follows the observation that, as there is no difference in the basic nerve units that are present at different electrode locations, one might expect that the slope of the psychophysical threshold function or difference limen would be constant at all electrode sites for a constant frequency.

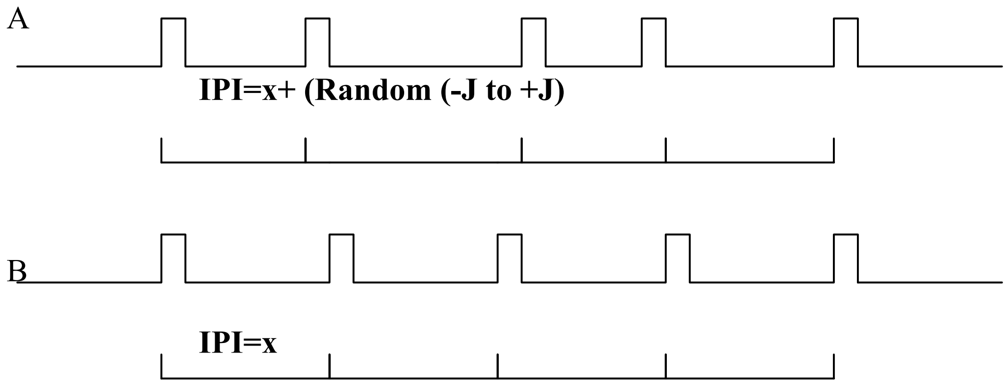


Figure 4-1 An example of a jittered pulse train (A) and unjittered pulse train (B). The jittered pulse train has the same mean IPI (x ms), but the pulse probability is a random distribution between –J_{max} and +J_{max} ms. The biphasic portion of the electrical pulse train has been omitted for simplicity.

4.11.2 Method

The author wrote purpose built software using Microsoft's Visual Basic programming language to provide a graphical user interface (GUI) so stimulation could be created and provided to the implant recipient.

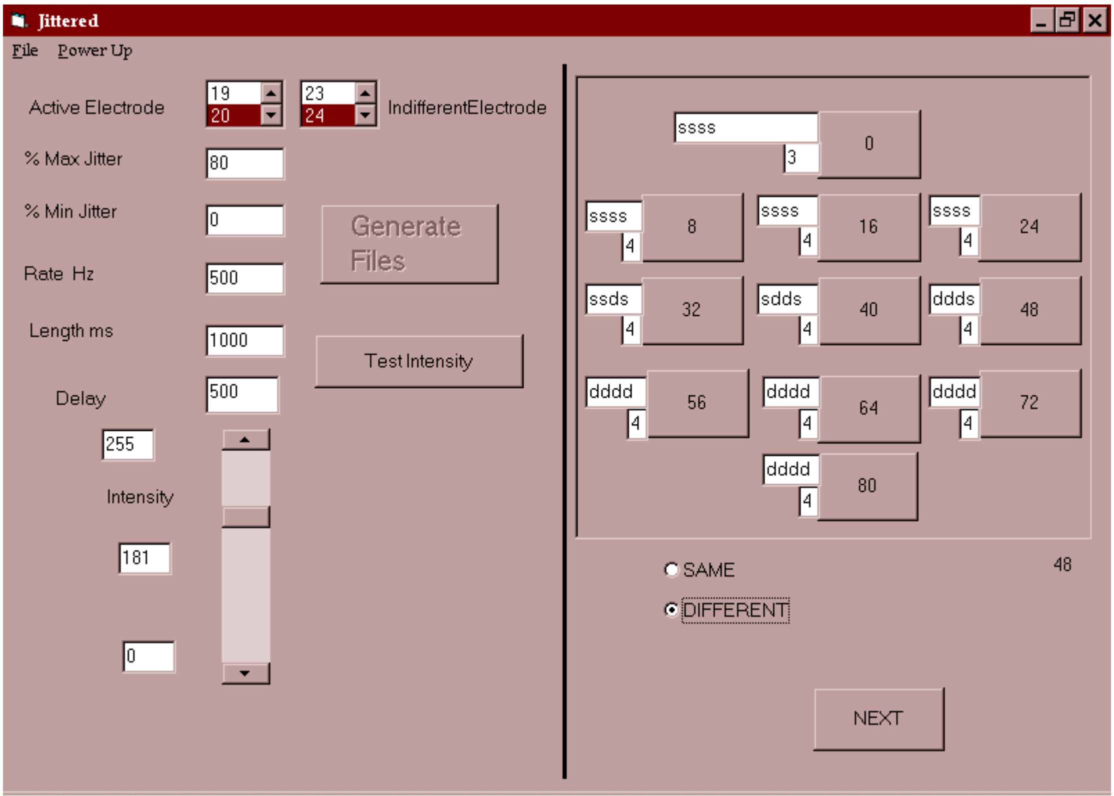


Figure 4-2 Graphical User Interface (GUI) of the jitter stimulation program.

The software allows selection of the active and indifferent electrodes via pull down lists. The main parameters are the numbers of the stimulating electrodes, the maximum jitter and stimulation rate. The length of the each stimulus and the delay between the stimulation pairs (ISI) is also selectable. These parameters default to 1000ms and 500ms respectively. To create a comfortable listening situation, the intensity is slowly increased until a comfortable level is found. All parameters are entered into the program via the GUI. Once all parameters are set and a comfortable level is found, eleven files containing the stimulation pulse information are created. These files contain increasing levels of jitter from 0% jitter up to the maximum percentage jitter level in equal steps of increasing jitter in each pulse trains. Stimulation was activated by clicking one of the eleven buttons on the screen using the mouse. This GUI communicates with purpose built software, the author also wrote, which was downloaded and running on the SPrint™ speech processor worn by the subject. By pressing the button, the file is downloaded to the processor and the subject receives the two pulse trains. One pulse train contained the selected level of jitter and the other pulse train has no jitter. The order of presentation was randomised each time the button is pressed.

After each stimulation pair, the subject responded by indicating if the two stimuli were the 'same' or 'different' (verbal response). The subject's response was entered by clicking on a checkbox on the screen and then clicking on the 'Next' button as can be seen in Figure 4-2. This response is displayed in a small text field next to the corresponding stimulation button so that the test instructor could see how many times each stimulation level had been played. The subject always faced away from the screen and could not see which pair was being played. Once each of the eleven stimulation pairs had been played at least four times, the electrode combination and/or stimulation rate are changed.

Each of the eleven stimulation sets were played a minimum of four times to the subject in randomised order giving a total of forty-four stimulations per electrode and rate combination. In this set of experiments, three different stimulation rates were used in combination with three different electrode combinations resulting in around 400 judgments enabling the thresholds, for low to high stimulation rates and low to high stimulation places within the cochlea, to be measured. The raw data were calculated as a percentage, correct for each level of jitter and converted to z scores using a standard z table. A least squares regression (linear) is used to find the line of best fit and from the calculated straight line, the point of subjective equality (PSE) and difference limen (DL) is found. To calculate the PSE, the point on the transformed psychometric function where z=0 is used and the DL is calculated by subtracting the value when the z score is ± 0.67 from the PSE value.

Five subjects took part in this experiment resulting in a total of 2000 judgments. All subjects are implanted with a Nucleus ™ 22-channel CI. Subject CIR2 used monopolar 1 (MP1) mode while the remaining subjects, CIR1, CIR5, CIR4, CIR3, used bipolar +2 (BP+2) mode.

4.11.3 Results

Individual subject data were analysed using analysis of variance (ANOVA) observing the factors of rate and electrode position. Subject CIR2's and CIR4's (see Figure 4-3) threshold for the detection of jitter is dependant on rate, significant at the 0.05 level. Performing further pairwise comparison using Tukey simultaneous tests reveal that CIR2's threshold for rate of the 100 pps condition is significantly lower ($p<0.05$) than the threshold for the rate of 500pps. No further significant effects were noted for individual subjects.

When the combined data was analysed it was noted that the jitter threshold was strongly dependent on rate, being significant at the 0.001 level, see Figure 4-3. Follow up Tukey tests suggest that discrimination thresholds for the condition of 100pps is significantly different from those for 250($p<0.05$) and 500pps ($p<0.0001$). Although the average measured jitter threshold is higher for the more basal electrode 5 than for electrode 12 and 20 (see Figure 4-4), the result was not significant ($p=0.056$). Figure 4-4 suggests that there is no strong overall trend with electrode position, although it is possible that there may be effects for individuals.

The jitter threshold levels from combined data is shown below

		CIR1	CIR2	CIR3	CIR4	CIR5	Mean	SD
Rate	100	38	32	28	22	24	29	5.8
	250	53	49	36	43	35	43	6.9
	500	73	61	37	45	44	52	13
		CIR1	**CIR2**	**CIR3**	**CIR4**	**CIR5**	**Mean**	**SD**
Electrode	5	72	51	39	33	41	47	14
	12	38	46	29	35	28	35	6.7
	20	54	44	31	42	35	41	7.7

Table 4.2 Jitter discrimination (%) for subjects for rate and electrode place. Jitter is always shown as a percentage.

		CIR1	CIR2	CIR3	CIR4	CIR5	Mean	SD
Rate	100	9.4	2.2	5.6	3.1	9.6	6.0	3.5
	250	6.7	5.0	4.4	2.7	4.5	4.7	1.4
	500	6.5	5.3	4.7	3.2	4.4	4.8	1.2
		CIR1	**CIR2**	**CIR3**	**CIR4**	**CIR5**	**Mean**	**SD**
Electrode	5	8.5	5.1	5.2	2.9	5.3	5.4	2.0
	12	7.6	0.9	7.6	2.7	16	6.2	5.8
	20	6.3	8	5.1	2.7	2.7	5.0	2.3

Table 4.3 Difference limens (%) for subjects for rate and electrode place.

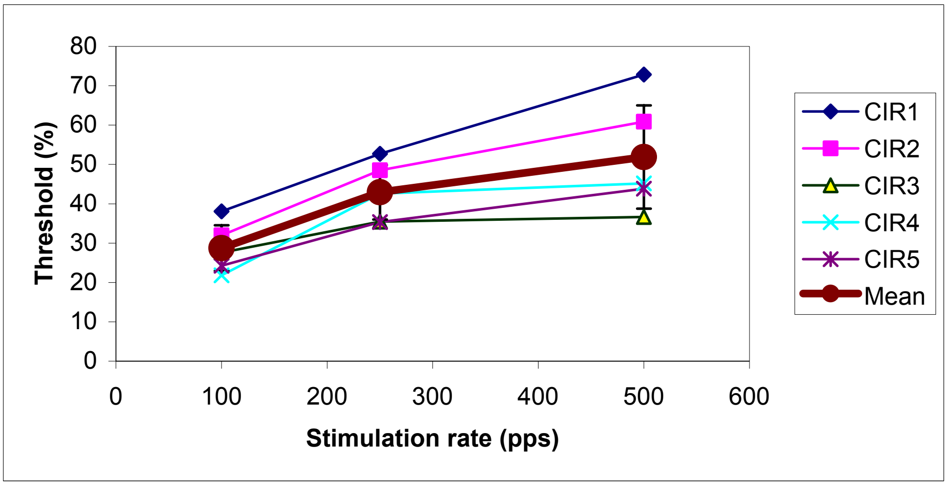


Figure 4-3 Jitter discrimination threshold as a function of stimulation rate

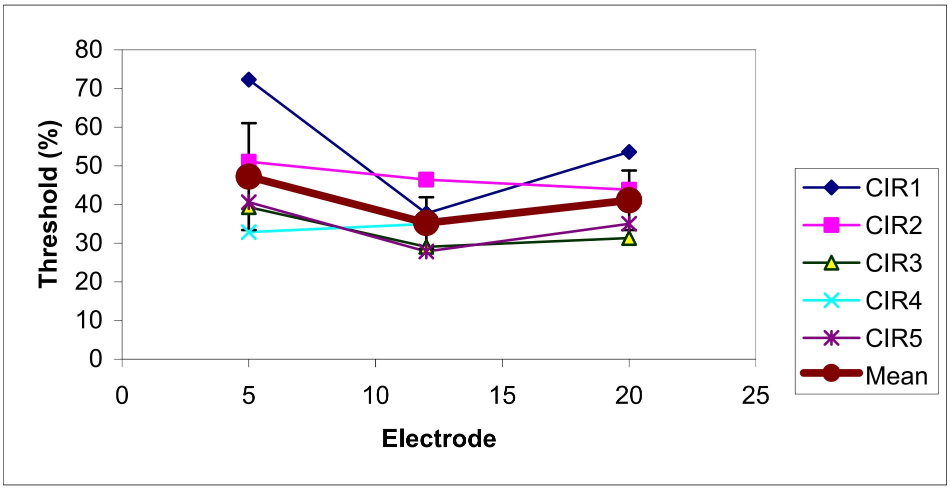


Figure 4-4 Jitter discrimination threshold as a function of electrode place

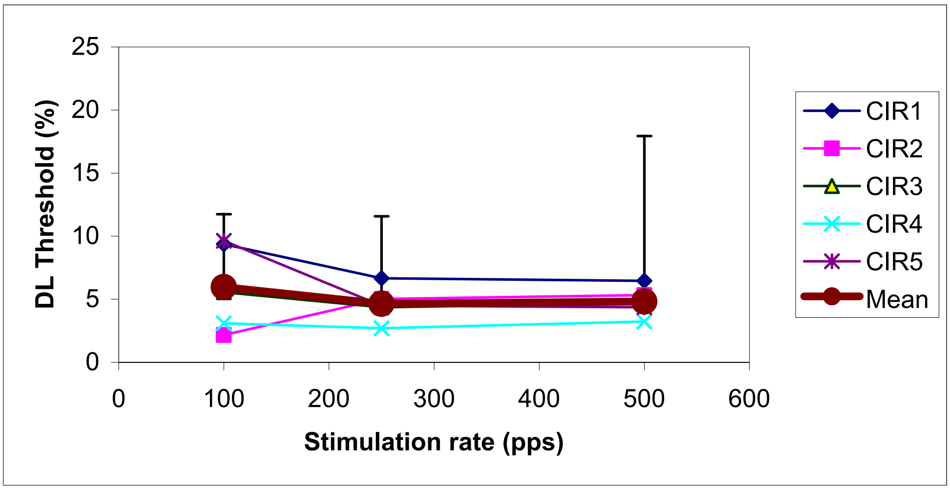


Figure 4-5 Difference limen of jitter discrimination threshold as a function of stimulation rate

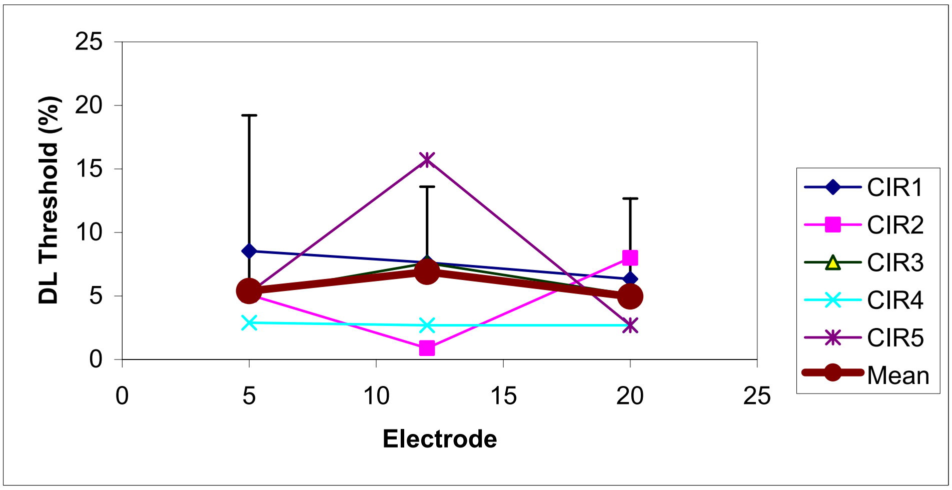


Figure 4-6 Difference limen of jitter discrimination threshold as a function of electrode place

When the difference limen is examined for its dependence on rate or electrode position no significant effects are found. Figure 4-5 shows the difference limen as a function of rate and Figure 4-6 shows difference limen as a function of electrode position.

Thus the conclusions on the hypotheses are:

i. Jitter threshold is dependent ($p<0.0001$) on rate of stimulation. The jitter thresholds are lower for low frequencies and become higher (i.e. jitter is harder to discriminate) as the frequency is increased.
ii. The jitter threshold is independent of electrode position.
iii. The difference limen or the slope of the transfer function is independent of rate.
iv. The difference limen is independent of electrode position.

Some subjects reported that the high frequencies and more basal electrodes sounded more ‘penetrating’ and that it was harder to pick the jitter in such conditions than for the lower pitch sounds (i.e. lower frequency and more apical electrodes).

4.11.4 Discussion

Current speech processing strategies such as SPEAK, ACE or CIS can jitter the stimulation rate, typically by 20%. This was historically done to reduce the ‘buzz’ produced by pure rate stimulation that created a pitch produced by stimulation rate so that place pitch information was not confused.
Comparing the results from this experiment with the results from the literature of NH jitter detection studies a dramatic difference between thresholds is observed. NH subjects can detect jitter as low as 0.1% whereas the lowest threshold a subject in this experiment could detect was a jitter at a level of 8%, a factor of 80 times larger. In this experiment there was a great deal of variation among subjects in their ability to detect jitter.
There are at least three possible sources for this difference between the results for NH listeners and CIRs. Firstly, the physiology that is typical of CIRs has some loss of ganglion cells and higher order neural pathways. This may affect the ability to convey presented information. A second problem is the method of presentation and nature of the signal. The electrical form of the signal is a ‘jittered in time’ pulse train, which superficially appears to be very similar in structure to an acoustically time jittered pulse

train. However, the electrical stimulation is presented to one electrode site that transfers a jittered time signal only. When the jittered signal is acoustically presented to the NH listener and analysed by the ear, the magnitude of the jitter will be represented both in the timing and spectral place of stimulation. Thirdly, the independence of the firing of nerves due to acoustic stimulation is much higher than that due to electrical stimulation. This may reduce the effective transmission of information that may be carried in discrete differences of timing of nerves.

Two of the subjects noted an interesting percept. CIR2 reported for electrode combination pairs 20/24 and 5/24, with a stimulation rate of 100pps, that the member of the stimulus pair with the jitter changes in sound quality during the presentation. This percept was present for levels of jitter at 36, 45 and 72%. The subject described the effect as “the signal starts off good then drops off in quality”. Subject CIR3 also noted this same perception for electrode combination 20/22 and 12/14 at 100pps for jitter levels 32% and 48%. This subject described the effect as going from “smooth to rough”. The subject reported that at 48% jitter, the signal changed from smooth to rough much earlier in the signal than when the jitter was at 32%. This subject perceived the same effect for electrode combinations 5/7 at 500pps for jitter levels 48% and 64%.

One possible cause of this percept is that the jittered signal, when delivered to the nerves, appears to be smooth as the jitter may be of the same order as the natural jitter and propagation variation times within the neural pathway. Perhaps the brain cannot detect that the signal is jittered and hence it is perceived as ‘smooth’ until a larger amount of the signal is sampled or integrated. As more of the signal is heard the higher order perceptual mechanisms then ‘decides’ that jitter is not natural and at this point the percept turns from rough to smooth. This possible explanation appears to be supported by the subject CIR3 who could detect that a signal with 48% jitter changed from rough to smooth at a much earlier stage in the presentation of the signal than the signal with 32% jitter. When the signal has a greater amount of jitter, it appears that a shorter time is required to detect the jitter. The author knows of no previous report suggesting that the length of presentation time is an important factor for perceiving jitter in a signal for a CIR. The author also knows of no previous observation of the ‘smooth to rough’ percept.

The signal processing strategies and stimulation techniques used by some CI signal processors require that the pulsatile stimulation be non-simultaneous, i.e. that electrical stimulation pulses can not be presented on different electrodes at the same time. To avoid simultaneous stimulation some pulses must be delayed or skipped. The question was how much delay or jitter could be allowed without having a perceptual degradation in the signal. These experiments have thus answered this question. Converting the lowest average threshold to time, the shortest discriminable average period variation is 1ms and it occurs when the maximum jitter threshold is around 50% for the 500pps signal. The smallest threshold detected in this experiment was when the maximum jitter in the signal was 0.7msec. This was when the threshold was 35% for the 500pps. The lowest threshold recorded for the 100pps signal was 8%, which correlates to a maximum jitter of 0.8msec that is of a similar order. These results suggest that, to keep within a non-detectable range, the maximum jitter allowed should be less than around 700μsec. These levels are within the capabilities of current speech processors and must be met for future music and speech-coding strategies to avoid a degradation of rate pitch and perceived quality.

To compare these data with those of Dobie and Dillier (1985) for two CIRs, the current threshold results must be converted to standard deviation of the IPI.

This Experiment		**Dobie and Dillier (1985)**
min level found	average level found	average level found for 2 subjects
5% (100Hz)	19% (100Hz)	17% (125Hz)
17% (250Hz)	29% (250Hz)	26% (250Hz)
23% (500Hz)	34% (500Hz)	16% (500Hz)

Table 4.4 Comparison of data of this experiment with that of Dobie and Dillier (1985).

There seems to be good basic agreement with the results of Dobie and Dillier except for the 500Hz condition. This experiment, however, has a larger number of CIRs and perhaps may be more representative of the general CI community. This experiment also:

i. Covers a much wider range of electrodes (places of stimulation) than the single channel implant used in the study by Dobie and Dillier (1985).
ii. Discovered definite trends for the dependence of jitter detection threshold on rate of stimulation.

4.11.5 Conclusion

The minimum threshold level for the maximum amount of jitter to be detected is around 700μs. This experiment also showed that the threshold is dependent on rate of stimulation but is independent of position (electrode). The difference limen of the threshold is independent of the rate of stimulation rate and electrode position.
This experiment supports previous experiments that jitter discrimination by CIRs is much poorer than with NH. When designing new signal processing strategies the limits found in this chapter will aid in determining how long a pulse can be delayed without resulting in a loss of sound quality.

Further experiments could explore the percepts noted by some subjects who described the rough to smooth change during the presentation of a jittered signal. Finding the limits of perception of detection of jitter as a function of the level of jitter and length of presentation could provide more insight into the neural mechanisms used to detect jitter.

Another future experiment could to non-simultaneously present jitter at multiple sites and also and at multiple frequencies to see if the threshold is affected.

Chapter 5
Pitch Pattern Perception

"Music produces a kind of pleasure which human nature cannot do without" - Confucius

5.1 Introduction

This study was designed to measure pitch discrimination when the test material was in a musical context, by measuring the audiated pitch that would complete a 'lead-in' phrase in a well-known melody.

The musical context of pitch discrimination is important:

> "The discrimination between two tones of different pitch without reference to keyality or tonality, and the discrimination between the two notes of different length without reference to tempo or meter, does not require musical thought" (Gordon, manual of IMMA, 1986).

The outcomes of this investigation will aid in the design of speech and music processing programs. If the results show that the accuracy in targeting the correct final note is low, then attention must be paid to improving the pitch discrimination ability. If the results show that accuracy is high but the ability to select the correct final target note is low, then attention must be focused to help re-map the correct pitch. Training may need to be given to help re-learn the interval sizes, target pitch and key.

5.2 Background

The brain appears to construct pitch at the auditory cortex. If the auditory cortex is damaged, as sometimes occurs with a stroke, the perception of pitch can be affected in such a way that only pure tone frequencies can be perceived and not complex pitches (Whitfield, 1980).

Neurophysiologists have found specific brain cells that process melodic contour i.e. the pattern of increasing and decreasing notes in music (Weinberger and McKenna, 1988).

Other cells have been found that seem to process specific harmonics (Sutter and Schreiner, 1991) while other cells in a specific part of the auditory cortex specifically process temporal and rhythmic aspects of sound (Hose et al, 1987). How music is perceived is generally regarded as a complex procedure that are processed by complex neural networks involving many parts of the brain.

A melody recognition exercise requires several response tasks. A discrimination of pitch change, direction and magnitude of change, and overall contour. In familiar melodies NH listeners rely on exact size of interval than overall contour. As pitch may not be well represented using conventional speech processing strategies, the exact interval sizes will most likely be poorly represented (Gfeller et al., in press).

Typical investigations with CIRs involve low numbers of subjects. This is due to the relatively low number of CIRs in any one location and the small percentage of these available and willing to take part in scientific investigations. To overcome this problem, this study was undertaken in conjunction with the University of Iowa implant center, in collaboration with the director of music therapy, Kate Gfeller. (A publication with joint authorship is planned).

5.3 Aims

The test used in this chapter is designed to measure pitch discrimination in a musical context. This is in contrast to other tests described in this thesis in Chapter 7 and 8 that provide isolated pitch discrimination techniques. Because of indications that different parts of the auditory system may be involved in the processing of melody and making isolated pitch comparisons, the measure of pitch discrimination found using this test will not be called a difference limen but the quantity measured will be called “dispersion” and is a measure of how consistently the participant can choose the ‘correct’ note in a melody.

The second product of this test is the measure of target pitch. The test presents the first phrase of a well-known melody where the final note is too low/high or correct. The participant, by selecting the appropriate response, causes the test to adapt to the target pitch they are audiating to produce a note to complete the phrase of the well-known melody. This is a measure of the pitch they remember as the correct note. For a CIR, the note selected to elicit the remembered pitch may be different from the accepted correct note, but it is perhaps a valid measure of the stimulation associated with the remembered note. The difference between accepted correct note and that targeted using the melody test will be referred to as 'error'. Several complications affect the choice of suitable stimuli:

If the correct final note was a repeat of a note in the preceding phrase it may result in a lower dispersion and error by helping the participant match a stimulus they have already heard.

If the final note of the phrase is higher than any of the notes of the lead phrase it may cause the subject to select a target note in the region of previously presented notes and hence result in a negative error.

If there are not enough presentations to portray the size of regular intervals of the equal tempered scale, the participant may not have a frame of key to correctly audiate the target note. This may be influenced by the compression the frequency has undergone to map it to the 25mm long electrode array.

5.4 Methods

The author wrote the test software using Microsoft's Visual Basic programming language. The testing software controlled a MIDI device that produced the audible notes. The testing software is self-administering and is able to be used by any participant capable of using a mouse to click the buttons on the computer screen. Upon starting the program a set of instructions is presented to the participant to instruct them what they are required to do. During the instructions, examples are given of the chosen melody with the correct note being much too high and much too low to demonstrate the purpose of the exercise. It was decided not to give the subject an example of the phrase with the correct last note as this may turn the investigation into a stimulus matching exercise instead of matching the correct note the participant has in their memory. A screen shot can be seen in Figure 5-1.

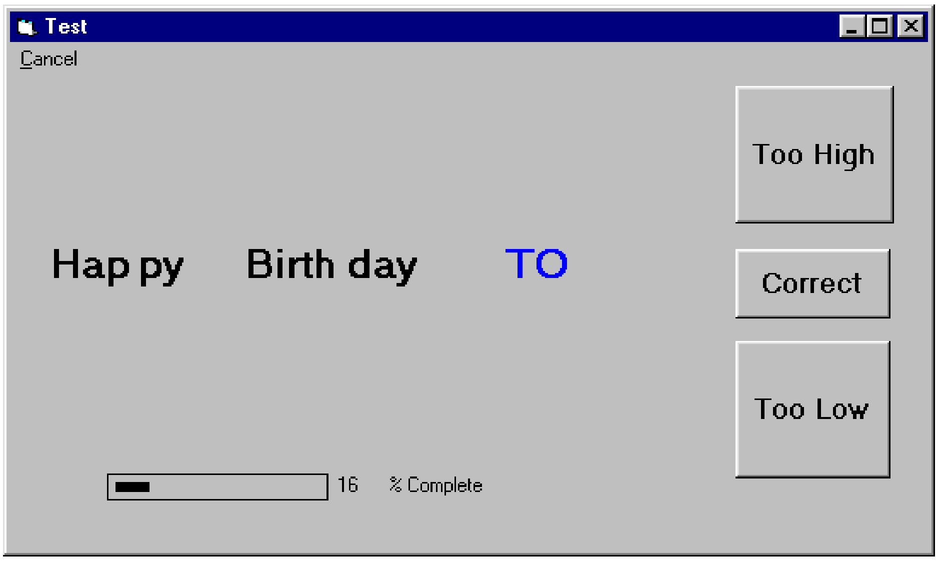


Figure 5-1 Screen shot from the melody testing software

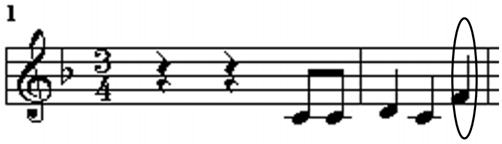
The testing software was originally written with three well know melodies that could be used. They were:

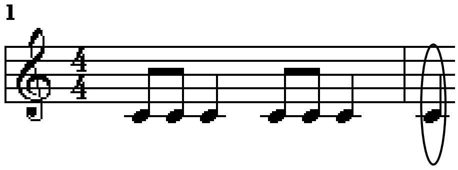
i. Happy Birthday (HB).

ii. Jingle Bells (JB).

iii. Three Blind Mice (TBM).

The notes that were presented in these tests are shown in Figure 5-2

i. Happy Birthday

ii. Jingle Bells



iii. Three Blind Mice

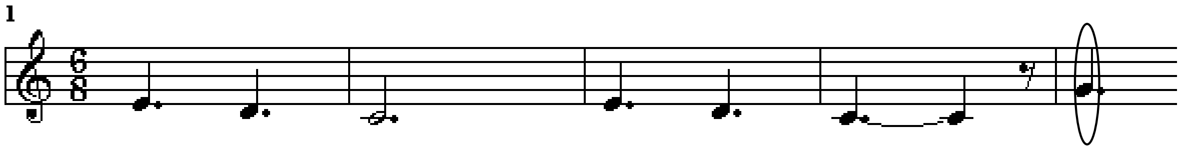


Figure 5-2 Musical notes presented in different versions of the tests. The last note of each phrase is the correct target note (circled), which the subject must find.

The contours of the three melodies provide an interesting contrast. Jingle Bells is the same note repeated and can be likened to a pitch-matching task presented in a melodic format. The Happy Birthday and Three Blind Mice melody provide some reference of interval size by presenting more than a repeated note. The Happy Birthday melody provides one note, two semitones above the other three. The target note is five semitones above the initial note. The Jingle Bells melody provides three different notes each two semitones apart. These three notes are then repeated. The target note is three semitones above the initial note. Comparing the presented range of each test, Jingle Bells is the smallest with a single note, Happy Birthday note has a range of two semitones, and Jingle Bells has the largest range of seven semitones. All tests were performed using a sampled piano stimulus. The piano stimulus is a complex harmonious wave, with varying spectral envelope and is easily identified as being musical by subjects. The tests were presented with the following options that were set as default but available from the pull down menus:

Timing:

i. Tempo of 90 beats per minute (bpm).

ii. A gap of 50ms between notes.

iii. No delay before the presentation of the final note.

Start Note:

iv. The first note in the phrase is middle C 261Hz.

Number of trials of one staircase:

v. Twenty i.e. forty presentations in all.

It was considered that there might be some benefit in inserting a delay before the last note sounded so that the subject would audiate the target pitch in place of the final note and when the pitch was produced the participant would compare these two. After pilot studies, it was decided, however, to make the melody sound as natural as possible, and thus there was no delay before the presentation of the final note.

5.4.1 Melody algorithm

The algorithm the testing software uses to adapt to the participants' responses is described below, using the example of the Happy Birthday melody to aid in the description.

The stimulus is the first five notes of "HAPPY BIRTHDAY" with the last note that goes with "TO" being too high, too low or correct. The aim of the experiment is to determine two main features. The first feature is the mean note that the implant recipient thought was the correct note for the "TO" in the initial phrase "HAP-PY BIRTH-DAY TO", from the tune "Happy Birthday". The measure of this is the mean error in semitones from the accepted correct note. The second feature, dispersion, is how confident they could repeatedly find that note. With two features being extracted from the test, four extreme conditions are possible:

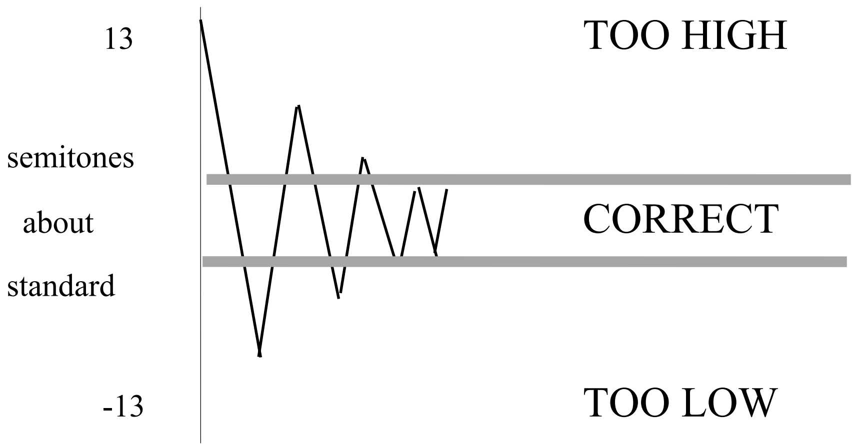
i. Low error – low dispersion: this indicates that they agree with the accepted hearing pitch of the note and that they find it easy to repeatedly find that note.
ii. Low error – high dispersion: this indicates that they agree with the accepted hearing pitch of the note but that they find it difficult to repeatedly indicate the direction of the 'correct' note.
iii. High error – low dispersion: this indicates that they disagree with the accepted hearing pitch of the note, which may be because of the mapping of intervals to electrodes. Alternatively they have fixed on a note which is incorrect due to poor pitch skills (as some hearing listeners do, e.g. picking the last note flat). When this note is found, however, they have good skills finding it again.
iv. High error – high dispersion: this indicates that they disagree with the accepted hearing for reasons discussed above, but they also have difficulty choosing a consistent note and maybe shows a 'guessing' technique was used.

The procedure chosen is designed to converge rapidly, be unpredictable by the participant and produce data that is relatively easy to analyse. The pitch of the last note is the only variable to be changed and is changed according to an adaptive algorithm. This algorithm, called 'random interleaved double staircase', has been adapted from Wetherill and Levitt (1965) and a summary of related documentation precedes this description.

The general features of the algorithm are:

i. There are two staircases randomly interleaved. One starts 13 semitones above, the other 13 semitones below the accepted 'correct note'.
ii. If there are fewer than 3 reversals, the step size is 4 semitones.
iii. If more than 3 reversals are recorded so far on the staircase, the step size is 1 semitone.
iv. If the response is 'too high', the tone is lowered by one step of the current size.
v. If the response is 'too low', the tone is raised by one step of the current size.
vi. If the response is 'correct', the tone changes in the same direction of the last change.
vii. A total of 20 trials from each staircase is presented (adjustable).
viii. If one staircase is completed first, the other is then sequentially completed.
ix. Analysable data are taken to follow the third reversal.

The staircase is designed to quickly approach the threshold region by taking large steps and once the threshold is found, smaller steps are used to gain a more accurate estimate. The assumed distribution is:



The error will represent the middle of the correct position and is calculated by subtracting the average of reversals after the third reversal from the real correct note in semitones.

The dispersion is calculated as the standard deviation of the points used to calculate the error.

5.4.2 Discussion of adaptive interleaved staircase technique

Many techniques have been adopted by researches for estimating points on a psychometric function. With the advent of computer control, rapidly converging adaptive procedures have been developed to gain accurate data in minimum time.

Cornsweet (1962) discusses one such technique, probably the simplest, being the adaptive staircase technique, *The Up and Down (UD) rule* (Dixon and Mood, 1948), or Bekesy audiometric method (Bekesy, 1947). If the subject responds with the parameter of the stimulus being too little; the parameter is increased upon next presentation and if the subject responds the parameter is too great; the parameter is reduced. The technique yields a 50% value or point of subjective equality (PSE) by computing the mean of the values after a given number of stimuli have been presented, representing the final level (Cornsweet, 1962).

Dixon and Massey (1983) suggest the staircase procedure has a saving over exhaustive measures by 30-40%, but is not good for estimating other points beside the 50% point.

Cornsweet illustrates there are four decisions to be made:

i. Where to start - it is suggested to start near the threshold to reduce trial number.
ii. How large the steps should be - no more than 2, 3 or 4 like responses before a reversal.
iii. When the series should stop - a compromise between large number of presentations and available time.
iv. When the series should be modified - large steps at beginning and smaller after the third reversal. (Cornsweet, 1962).

The advantage of the staircase method is that it is very efficient as most stimuli are near threshold. The disadvantages are that the subject is aware of adaptive presentation and may guess through a pattern of responses that may be interpreted as meaningful data. As

the judgments become more difficult, the procedure may also become distressing (Cornsweet, 1962).
The former of the disadvantages may be overcome by randomly interleaving a second staircase. With this, the subject feels no constraints of single staircase and it is not possible to give meaningful data by guessing or by following a pattern of response. It also avoids sequential interaction (Treutwein, 1995). If both staircases converge, the subject must be responding to the same aspect of stimulation (Cornsweet, 1962).

The two features we are attempting to extract are the error and the dispersion. These can be calculated in several ways. Treutwein (1995) calculates the mean by averaging the reversal points (called a mid-run estimate). Cornsweet (1962) calculates the mean of all values after a given number of stimuli when the series has leveled off. Kollmeir et al. (1988) estimates the mean by averaging levels after third reversal to the end of an even number of runs. Wetherill and Levitt (1965) average all the peaks and valleys to estimate the mean.

The more difficult feature to extract is the dispersion. We are mainly interested in an accuracy value that estimates the variance of the mean error for subject comparison. The dispersion calculation is difficult as successive peaks and valleys are not statistically independent. Wetherill and Levitt (1965) suggest that the correlation between the averages of pairs of runs is much smaller than the total set of runs and an approximate standard deviation can be obtained by calculating the variance of the average of peaks and valleys in pairs of runs. Dixon and Mood (1948) provide a method for obtaining an estimate for the standard deviation if the step size is around the size of the standard deviation. This prior knowledge is unavailable for this experiment, but could be obtained with a pilot study. To avoid a dispute of technique in calculating the variance, an error (not standard deviation or variance) that represents the dispersion of data will be used. The dispersion is calculated by incorporating all the points after the third reversal in an even number of runs and calculating their standard deviation. This value is not an ideal measure of the standard deviation because of the discrete nature of the steps. The measures sample a continuum at intervals that, far from being continuous, are possibly larger than the standard deviation of the underlying function. Further they represent the results of a 3 way choice (high, low and correct). Nevertheless, this value can then be consistently used to compare subjects and the changes in accuracy in producing a target note should be reflected in changes in this value.

5.5 Participants

Twenty-five CIRs took part in this study. Twenty-three participated in the "Happy Birthday" version of the test, thirteen participated in the "Jingle Bells" version of the test. Twelve participated in both the "Happy Birthday" and "Jingle Bells". Three participants trialled the "three blind mice" version but later, this version was discontinued. Twenty-one NH participants also took part in this study for comparison. Twenty participated in the "Happy Birthday" version of the test and nine participated in the "Jingle Bells". Nine participated in both the "Happy Birthday" and "Jingle Bells" version of the tests.

There was a range of brand of implants used as well as speech processing programs used. This study was intended to be a general comparison between CIR and NH groups and not intended to compare implants and speech processing strategies. A CI participant description can be found in Table 5.1.

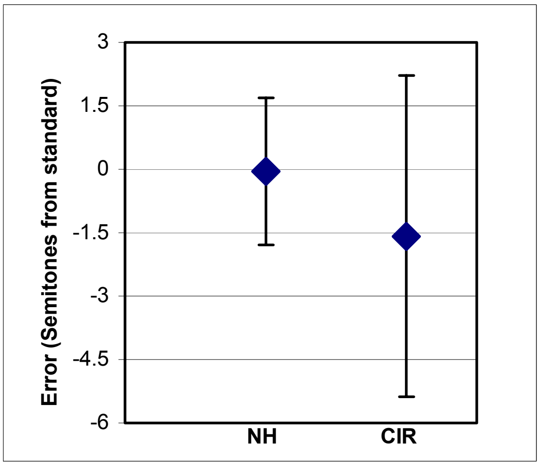
Subject	Implant	Strategy	Age	Sex	Usage(mo)
1	NUCLEUS	MSP/SPEAK	51	M	118/0
2	CLARION	SPEAK	65	M	60
3	CLARION	CIS	61	M	36
4	CLARION	CIS	34	F	36
5	INERAID	ANALOG	59	F	108
6	INERAID	ANALOG	75	M	126
7	MED-EL	CIS	61	F	126
8	NUCLEUS	SPEAK	48	M	112
9	MED-EL	CIS	69	F	126
10	NUCLEUS	SPEAK	72	F	106
11	MED-EL	ANALOG	44	M	44
12	NUCLEUS	SPEAK	73	F	73
13	NUCLEUS	SPEAK	38	M	38
14	NUCLEUS	SPEAK	52	F	52
15	CLARION	CIS	41	F	41
16	MED-EL	CIS	78	F	78
17	MED-EL	SPEAK	69	F	69
18	NUCLEUS	SPEAK	63	M	63
19	NUCLEUS	SPEAK	74	F	74
20	MED-EL	CIS	80	M	80
21	NUCLEUS	SPEAK	78	M	78
22	CLARION	CIS	51	F	51
23	NUCLEUS	ACE	63	F	50
24	NUCLEUS	SPEAK	40	M	12
25	NUCLEUS	ACE	62	M	24

Table 5.1 CI participant details

The NH group ranged in age from 14 to 55 (mean 23.5), 88% were female and 56% were involved in music in some way but were not professional musicians.

5.6 Results

It appears that the NH group is very accurate with a mean error of almost zero (–0.048 semitones). Most NH participants score a perfect zero. The CIR group, however, when averaged across all tests have a larger error of –1.5 semitones. The sign of the error indicates that the targeted pitch was below the accepted value. The standard deviation of the error is also larger for the CIR group, 3.8 semitones compared with 1.7 semitones for the NH group. This is not a surprisingly poor score considering the reduced pitch information available to the CIR using traditional speech coding strategies.



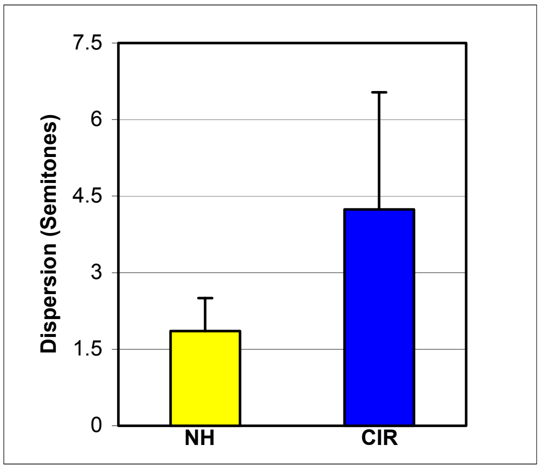


Figure 5-3 Combined error and dispersion of all 3 tests for NH and CIR groups.

Consider now just the CIR group and compare the three versions of the test. The measured error is shown in Figure 5-4. The target pitch, for the ‘Happy Birthday’ test, was 3 semitones below the accepted pitch, which is the same note as already heard in three out of the four lead in phrase notes. The targeted pitch of the ‘Jingle Bells’ test was around half a semitone above the accepted pitch. The targeted pitch of the ‘Three Blind Mice’ test is around one and a half semitones above the accepted correct pitch.

Examining the error of the NH group reveals little difference among the error, for the three tests. It is interesting to note that the trend of the sign of the error matches that of the CIR group with the 'Happy Birthday' test scored below the target pitch, the 'Jingle Bells' test reported accurately and the 'Three Blind Mice' test reported above the correct pitch. The standard deviation of the error for the NH group, however, is much larger for the 'Happy Birthday' test than for the 'Jingle Bells' test as can be seen in Figure 5-4.

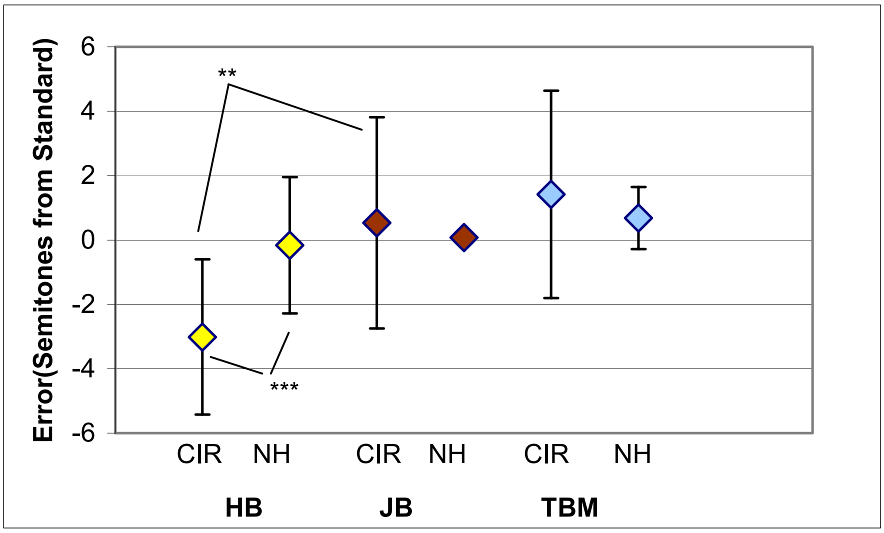


Figure 5-4 Error of the different tests for the CIR and NH groups. The difference between the CIR and NH Happy Birthday scores, significant at the 0.001 level (p=0.001). The difference between CIR, Happy Birthday and Jingle Bell scores is significant at the 0.01 (p=0.0095).

When examining the dispersion measure for the different tests, for the CIR group shown in Figure 5-5, there does appear to be a difference between the 'Happy Birthday' test and the other two tests. The 'Happy Birthday' test produces a much higher dispersion of around five semitones compared to around three semitones for the other two tests.

The dispersion of the NH group appears to be reasonably consistent across all three tests suggesting that NH subjects have consistent ability to target repeatedly the same pitch.

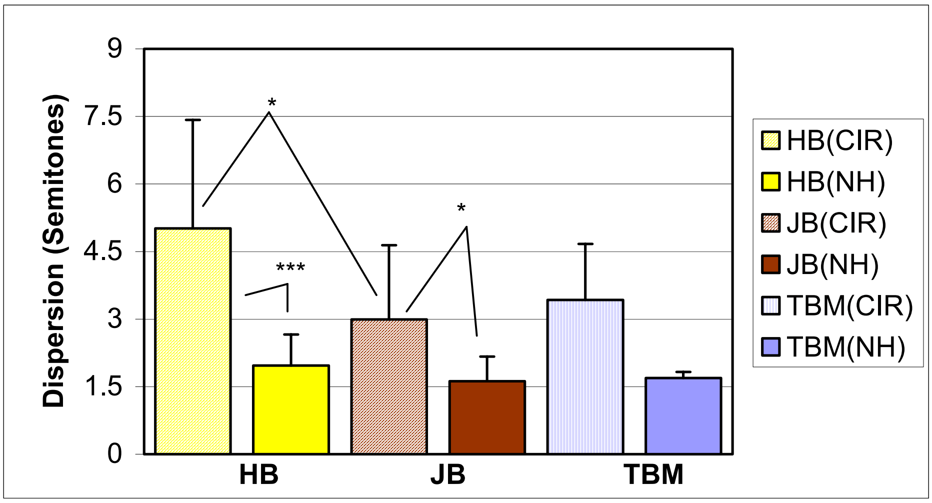


Figure 5-5 Dispersion measure for three different tests for the CIR and NH groups. The difference between the dispersion measures of the CIR Happy Birthday and Jingle Bells test were significant at the 0.05 level (p=0.0256). There is also a difference between the dispersion measures of the NH and CIR groups significant at the 0.001 level (p=0.000) for the Happy Birthday test and 0.05 level (p=0.026) for the Jingle Bells test.

5.6.1 Correlations

Correlations were performed on the data observing the error and dispersion due to the effects of test type, CIR or NH groups separately and combined. It was found that the dispersion was significantly correlated between the JB and HB tests for the combined data of the CIR and NH groups. Correlation is significantly different from 0, at the 0.05 level (p=0.029), at 0.487.

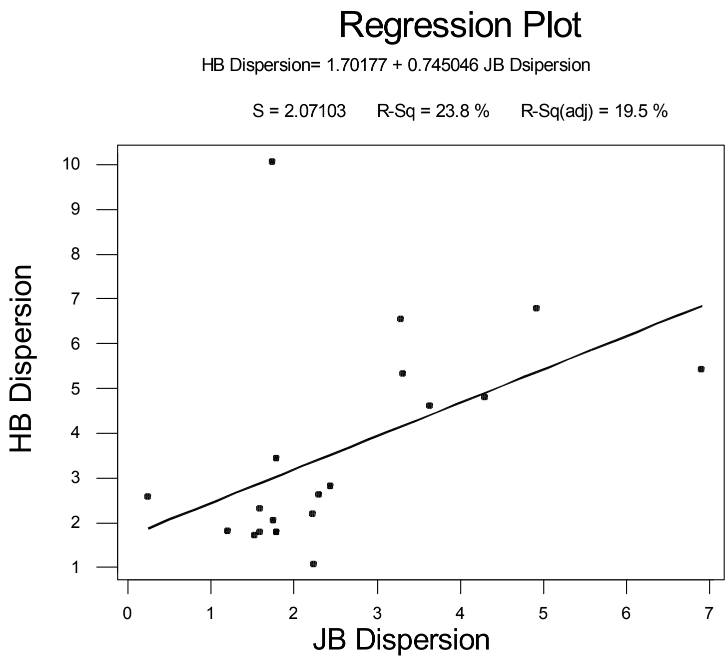


Figure 5-6 Regression plot of combined data of NH and CIR groups comparing dispersion on HB and JB tests. There is a significant positive correlation (0.05 level) between the dispersion on these two tests.

ANOVA's were performed on the data observing the error and accuracy due to the effects of implant type, strategy and sex. Correlations were calculated on the data measuring the interrelation of error and dispersion with age and usage.

A correlation was found between error and age when subjects participated in 'Jingle Bells' test. The Pearson's product moment correlation coefficient was 0.63, significant at the 0.05 (p=0.015) level. There was a correlation between the age of the subject and the error of selecting the correct interval, as can be see in Figure 5-7. The younger subjects target a pitch too low while the older subjects target a pitch too high, with a correct target pitch for the subject aged around 60. This effect cannot be immediately explained but may possibly related to nerve survival effecting shifts in pitch perception, or it may be a Type I error.

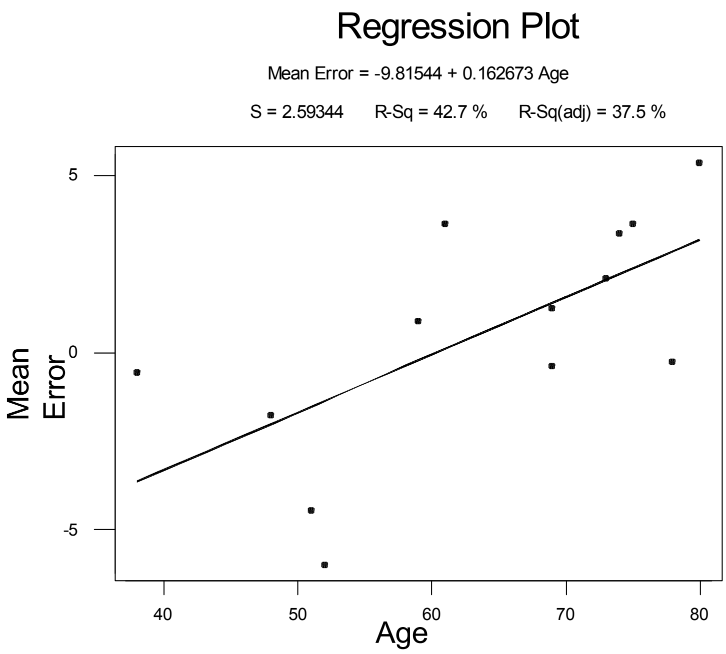


Figure 5-7 Regression of Error and Age for the Jingle Bells test (CIR group)

5.7 Discussion

In the 'Jingle Bells' test, the final target note has the same pitch as all preceding notes, but this test could be said to have minimal melodic content and thus reduces to a 'match the note' exercise. Some subjects reported that they knew the final note was the same note as all the lead in notes and they just had to 'find' the matching note.

The 'Happy Birthday' test results for CIRs had, on average, a negative error. However, there is a positive error for the few subjects who trialled the 'Three Blind Mice' version of the test, even though the HB and TBM versions had a target note higher than all lead-in notes.

In the example it appears that, if the correct final note is one of the presented notes, there may be a higher chance that the note chosen is one of the presented notes in the lead in phrase. The target pitch chosen by the CIR's for the HB test was 3 semitones lower than accepted correct note. This note however is the same as the third note of the four note lead in phrase which appears only once in the lead in phrase along with 3 repetitions of a lower note. Played to a NH participant, the phrase with the chosen note is harmonious, but not melodically 'correct'.

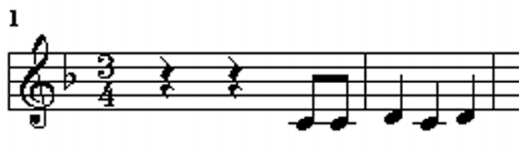



Figure 5-8 The Happy Birthday test phrase with average last note as judged to be the correct by the majority of CIRs.

5.8 Conclusion

Two measures of pitch discrimination in a musical context were obtained: one is of the error, or the difference between the subject's target note and the accepted correct note, the second measure obtained was the dispersion, which measured the variance, or confidence the participant has in finding their chosen target pitch.

Both CIR and NH groups performed the 'Jingle Bells' test the best, with the accuracy of the NH group being greater then the CIR group.

The HB test appears to be judged differently by the NH and CIR group. The CIR group not only showed lower accuracy, but also judged the correct final note 3 semitones too low. There may be several reasons for this. For example, they could be trying to match a note that was already in the lead in phrase. It is also possible that the participant had rarely heard the HB tune played/sung in tune and had difficulty remembering the correct note. Over the course of this thesis the author has paid unusual interest to the intonation of the singing of Happy Birthday in social situations. The author proposes that the implant recipient may have added difficulty to ascertain the correct notes if such social situations were the only presentation and training undertaken. It is also possible that the ability to audiate a single note while many incorrect versions are presented may be a difficult task.

An unexpected finding in the chapter has been the inaccuracy of at least some NH listeners, indicated by the both the standard deviation of the error and the dispersion measure. However the dispersion of the CIRs are still far larger than that of the NH participants.

Tests of this type are more difficult than tests where same or different judgments or pitch direction judgments are made when comparing two isolated tones. This test may engage a part of the brain, which would be used in judging the melodic presentation as a whole as opposed to isolated tones.

An important measure introduced in this chapter is dispersion, which is a measure of accuracy as it provides a type of pitch discrimination for audiation i.e. accuracy for judging a pitch heard internally compared with an audible pitch.

In conclusion it was decided that the 'Happy Birthday' would be used to measure dispersion and error, and the 'Jingle Bells' test could be used as an alternate to a pitch matching procedure, like the PDT described in chapter 8. Electrodograms of a piano scale and a presentation of the Happy Birthday test are shown in Appendix A.

Chapter 6
Tonal Language

6.1 Introduction

This chapter aims to investigate the suitability of present CI speech processing designs for use with tonal languages i.e., languages in which tone carries lexical information. Previous studies will be described and a summary and interpretation of these results will be discussed.

The experimental study that was carried out in this chapter aimed to determine, through simulations on NH listeners how parameters important in a tonal language were perceived after processing by a CI processor. Studies involving CIRs who speak a tonal language are quite difficult to conduct due to the small number of subjects there would be in Australia that have a CI and speak a tonal language. The results suggest that some current processing schemes do not provide the pitch resolution necessary to distinguish important tonal features of a tonal language. New processing schemes will be described that may improve the perception of a tonal language by a CIR.

There are limitations to the use of simulations as models of CIRs. However, if particular types of information are removed in the experiment, the simulations are useful in modeling how well one can analyse a signal without that information.

The pilot study described in this chapter was presented as a poster at the International Conference on Spoken Language Processing (Sydney, 1998) and was published as part of the conference proceedings (Fearn, 1998).

6.2 Non-tonal language design

The design and initial distribution of the early CIs was in Western countries, particularly Australia, North America and parts of Europe. These countries have remained the market for which much of the design, manufacture and distribution of CIs has been aimed. Naturally, these devices were initially developed to process the phonemes of non-tonal languages. Thus, the speech processing strategies that have been developed to drive the implants provide detailed information about the spectral envelope and transients, the information necessary to identify the phonemes of English and other non-tonal languages. In recent years, CIs have been rapidly distributed throughout East Asian countries (Hiki et al., 1998) such as China, Vietnam, Malaysia, Thailand etc. In a large proportion of these countries tonal languages are spoken. These are languages that carry information in the tone, as well as in the spectral envelope and transients. A ‘tone’ can be described as a distinct level of pitch or a change in pitch throughout the duration of the word. In non-tonal languages, intonation is used to convey such information as whether a phrase is a question (in English, by rising pitch), or a statement (by falling pitch). In tonal languages, single words or syllable that are otherwise phonetically identical can be distinguished by the tone.

One quarter of the world’s population speaks Chinese. Although the written language is invariant, there are many different spoken Chinese languages; two of the most important being Mandarin and Cantonese. Mandarin is spoken primarily in Mainland China, especially in the north. Cantonese is spoken primarily in the south of China and in other Chinese settled countries like Malaysia and Hong Kong (Comrie, 1987).

Other languages that use tonal features to carry lexical information are Vietnamese, which uses 6 tones, Thai that uses 5 tones and some dialects of Korean. Other Korean dialects, like some African languages, including Swahili, have evolved and have almost completely changed from tonal to non-tonal (Cummins, 1998).

Most previous studies using multichannel CIs use non-tonal languages to evaluate speech and any auditory improvements (Huang et al., 1996). It is of much interest to researchers to compare speech perception of tonal languages like Chinese and non-tonal languages like English to ascertain how far a formant based speech coding strategy is applicable to both languages (Xu et al., 1987). The CIs that were designed for non-tonal languages did not incorporate the ability to discriminate small changes in pitch in their engineering. In some Asian countries, the success of the cochlear prosthesis may partly depend on this capacity to transmit these characteristics (Huang at al, 1995).

Studies have been carried out in improving the filterbank frequency boundaries for non-English, non-tonal languages such as Spanish (Aronson & Arauz, 1995). Filter boundaries were adjusted to better differentiate the 5 vowels of Spanish Rioplastenese so that vowels with similar formant positions could be identified more easily. This technique significantly improved vowels and bisyllable discrimination by 11% (73% to 84%) and 8% (64% to 72%) respectively. Other studies have investigated tonal languages as perceived by CIRs (Zeng, 1999b; Kwok et al., 1991; Huang et al., 1996; Wang and Huang, 1988; Huang at al, 1995; Xu et al., 1987; Tang et al., 1990). However, none of these studies have altered or tried to optimise speech-processing parameters to improve the perception of tonal languages. No studies have been conducted with pitch-accented languages such as Swedish or Japanese.

6.3 Tonal languages

Mandarin officially uses four different tones and Cantonese six. For a typical male speaker of Mandarin who uses an F0 in the range of 88 to 190Hz, the four tones can be illustrated with the example of the syllable transcribed as /ma/:

i. Tone 1: Flat F0 contour with no inflection in the range of 110Hz and 150Hz, /ma/ meaning mother.
ii. Tone 2: Rising inflection between 100Hz and 140Hz, /ma/ meaning hemp.
iii. Tone 3: Falling then rising between 90Hz and 116H, /ma/ meaning horse.
iv. Tone 4: Falling inflection between 150 and 108Hz, /ma/ meaning scold.

Ciocca et al. (2000) suggest that changes in amplitude patterns and duration have been found to be reliable acoustic cues that can differentiate words differing primarily in tone in Mandarin.

The 6 tones of Cantonese can be illustrated using the example of the syllable transcribed as /si/:

i. Tone 1: High Level/Falling, /si/ meaning poem.
ii. Tone 2: High Rising, /si/ meaning history.
iii. Tone 3: Middle Level, /si/ meaning test.
iv. Tone 4: Low Rising, /si/ meaning time.
v. Tone 5: Low Falling, /si/ meaning city.
vi. Tone 6: Low Level, /si/ meaning event.

There are also three ‘entering tones’ that are used exclusively with syllables with stop consonants /p t k/ and are short forms of tones 1, 2 and 3 (Kwok et al., 1991). The most common tones are high level, high rising and low falling (Barry et al., 2000).
In Cantonese, the smallest meaningful unit is the syllable and can be a vowel (V), consonant vowel (CV), consonant vowel consonant (CVC) or vowel consonant (VC). 98 % of the Cantonese syllables are in the form of CV or CVC (Kwok et al., 1991).

Wong (1984) cited a collection of 573 syllables having 10 000 different meaning and Bai (1984) cited 409 syllables based on 30 000 character entries. Tones have a heavy functional load in the transfer of lexical information (Yuen, 1984) by further dividing the relatively small number of syllables into different meanings. The tone is very important when the syllable is isolated but the importance of tones is lessened with contextual queues (Kwok et al., 1991). However, in a noisy situation, the tone is an important factor as it adds extra information to understand the sentence (Yuen, 1984). Cantonese children are thought to acquire tones early in language production because pitch is the feature of speech most salient to a child (Barry et al., 2000). However, children are thought not to be able to recognise isolated lexical tones until around 4 years of age (Hiki et al., 1998).

Thai is spoken by around sixty million people and is primarily a monosyllabic, tonal language. There are 44 consonant characters in Thai, representing 20 consonant sounds. The 44 consonants are grouped in 3 classes HIGH, MID and LOW classified by their tonal quality. There are 9 short monophthongs (single) vowel sounds and 9 long counterpart vowel sounds plus 3 diphthongs sounds in Thai, together with 15 basic vowel-characters. There are 5 tones in standard Thai: mid (or monotone), low, falling, high and rising. There are 4 tone marks in Thai. They are placed above the initial consonant of the syllable whose tone they mark. The tonal value of the tone marks depends on the class of the initial consonant and the type of syllable that they mark. The 5th unmarked tone takes the default value as determined by the tone class.

In these studies Thai and Cantonese languages were chosen because:

i. Unlike Mandarin, the tonal features of Cantonese are without changes in amplitude and duration, which can be used to predict the correct tone.
ii. They have a high number of tones, six, reducing the chance level of identification to 16.67%.
iii. Several native speaking subjects were available to test.

6.4 Implants used with testing tonal languages

Two classes of CIs have been used in previous studies that investigated perception of tonal languages. These are the 3M/HEI (House Ear Institute) single channel device, (Wang and Huang, 1988; Kwok et al., 1991; Tang et al., 1990) and the Nucleus CI22 multichannel device (Huang et al., 1995; Huang et al., 1996; Xu et al., 1987; Barry, 2000). One study, (Ciocca et al., 2000) used the Nucleus CI24 multichannel device, which is a newer version of the Nucleus CI22. The studies that used the CI22 device used a speech processing strategy (Multipeak) that transmits formant information F1 and F2, using amplitudes from each respective formant, A1 and A2, as well as three additional bands of high frequency spectral information at a pulsatile stimulation rate of F0. The study that used the CI24 device (Ciocca et al., 2000) used a speech processing strategy (SPEAK) that performs an FFT on the speech signal and stimulates channels of maximum energy via pulsatile stimulation at a rate of 250Hz per channel. It should be noted that the multichannel Multipeak strategy uses stimulation at the voice pitch rate. The 3M/HEI device, a single channel implant stimulates the nerves using electrical analogue stimulation at the corresponding voice pitch rate. The SPEAK strategy delivers voice pitch via a different method that is primarily by choice of electrode (place pitch) and via temporal modulations of the 250Hz stimulation rate. One recent study (Zeng, 1999b) using four CIRs who were native Mandarin speakers and who used the Multipeak speech processing strategy (which presents the F0 voice pitch rate), examined rate discrimination and tone discrimination. It was found that rate discrimination and tone recognition, while varying greatly amongst individual subjects, were highly correlated, which suggests that good rate discrimination ability is required to be able to differentiate tonal variation in that language.

No studies were found that investigated tonal perception by CIRs who spoke Thai.

6.5 Testing techniques

Previous studies have used various methods of testing speech perception of tonal languages. Huang et al. (1995) used Mandarin tone perception tests which included ten items of four forced choice words whose phonological structure were identical except for different tonal features. This was done to compare preoperative recognition with hearing aid with postoperative recognition with a CI. Kwok et al. (1991) tested tone recognition ability by using the syllable /fu/ which has six tones in a 6 by 6 confusion matrix. Each word could mean “husband”, “to hold”, “tiger”, “woman”, “trousers” or “father”. Ciocca et al. (2000) recorded the Cantonese syllable /ji/ in the phrase “I will read ___ for you to hear” so that the tonal words are in the medial position to prevent change in intonation that may affect perceived tones. They were isolated and presented as pairs in 8 tonal contrasts, each contrast was presented four times in each block, and each subject listened to five blocks. Perception was tested via a picture identification test. Barry et al. (2000) investigated speech production by monitoring F0 as a measure for acquisition of tonal and vowel learning.

From these studies, it appears that the best method is to excise the stimuli medially of a sentence and present them in an isolated fashion randomly to the subject for identification, while displaying the total set for the subject to choose from.

6.6 Summary of results from previous studies

A study using the single channel device (Huang et al., 1995) suggested a significant increase from 35% correct tone recognition preoperatively with hearing aids to 68% tone recognition postoperatively with CIs. One subject scored as high as 90% (chance level was 25%). Scores on other speech tests also improved by a comparable amount. It may be that the acoustic cues of the F0 patterns in the four mandarin tones are extracted by Multipeak and when presented at differences in rate of stimulation, are perceived as change in pitch. Kwok et al. (1991) in a study with eight subjects, used the six tones and subjects scored an average of 38% where the chance score was 17%. One subject achieved 65% while the lowest scoring subject was at chance level. All subjects, but one, scored higher than hearing aid subjects, suggesting that the single channel device can transmit pitch change and stress like features of the auditory signal. In a study with one subject using a Chinese segmental tonal test the subject scored a surprisingly high 100% (33/33). This suggests that pulse rate was adequate for correct tone identification as this subject used the F0F1F2 strategy that stimulates at the rate of F0 (Xu et al., 1987). Tang et al. (1990) in a study with four subjects and using six tones (17% chance), found subjects averaged 31% correct identification with results ranging from 27% to 47%. All but one subject scored higher than subjects who used conventional hearing aids. In a study using the SPEAK speech processing strategy, Ciocca et al. (2000), tested subjects with tonal comparisons and found that there were no significant results for the tonal contrasts. The average score was between 48 to 58% (chance 50%). Only one subject scored above chance on one of the eight contrast tones. One possible reason for this low identification rate may be due to the SPEAK strategy not extracted effectively F0 which may assist in the identification of the word. The low scores on these tests may be due to subjects being children CIRs whereas studies of Huang (1995) whose subjects scored an average of 68% from the four Mandarin toned the subjects were postlingually deafened adult users.

As explained in the introduction, a multichannel CI consists of an array of electrodes, up to 22, inserted into the scala tympani in the cochlea. These 22 electrodes are stimulated by passing current between them to activate the surviving nerves in the organ of Corti or more possibly the spiral ganglion cells in the modiolus. A typical healthy cochlea may have 30 000 nerve cells stimulated by the motion of the basilar membrane. A typical CI user may have only 10 000 surviving nerve cells (Clark, 1992) stimulated electrically, by at most, 22 electrodes. As a result, a large reduction in spatially discriminable stimuli would be expected. There exists however, the temporal component of coding the input signal that may accurately convey 'pitch' up to 800 Hz (Pijl and Schwarz, 1995; Fearn and Wolfe, 2000).

Some recent speech processing strategies do not attempt to extract and present F0 information to the CIR by rate of stimulation. These strategies choose an arbitrary rate and rely on the tonotopic arrangement of the cochlea signal as well as amplitude modulation to convey the pitch of the incoming signal (Whitford et al., 1995). (See Clark (1992) and Blamey et al. (1985) for a complete description of recent strategies and results achieved). These recent strategies that present detailed spectral information have resulted in speech scores better than those strategies that pick formants. The reasonably good tone perception results in studies using single channel device or the CI22 device and the Mulitpeak strategy compared with the chance levels of tests using the CI24 device using SPEAK seems likely to be explained by the difference in the way F0 is presented. The SPEAK strategy, however, does show an improved overall speech score when compared to the earlier strategies that presented F0.

6.7 Simulations

Several researchers have experimented with simulations of CI speech processing strategies to study the effects of varying parameters on speech perception when presented to NH listeners. Dorman et al. (1997) and Loizou et al. (2000) examined speech intelligibility as the number of simulated 'channels' was varied. Dorman et al. (1997)

simulated changes to speech due to variation in insertion depth of the electrode. Shannon et al. (1995) examined a particular case; the adaptation to the spectral shift resulting from simulation of an electrode array that was not fully inserted. Rosen et al. (1997) examined the importance of temporal cues on speech and Shannon et al., (1998) measured speech perception with altered spectral envelopes. Loizou et al. (1999) examined recognition of sentences in noise while varying the number of total channels and also the number of channels selected each processing cycle. In one investigation that examined tonal language with simulations Fu et al. (1998) did not necessarily aim to simulate CI speech processing but that divided speech into one to four frequency bands. The simulation extracted the amplitude envelope in each band. The amplitude was then used to modulate noise of the same bandwidth as that of the analysis band. Using Mandarin and English with twelve subjects, vowel, consonant and sentence recognition improved monotonically with the number of bands presented to the subject, but tones were identified at around 80% correct regardless of the number of bands.

The study reported in this chapter investigates two different strategies by simulation with NH listeners. One strategy utilises the spatial, or tonotopic, arrangement of the cochlea and presents stimulation at an arbitrary rate. The other strategy uses the spatial tonotopic arrangement of the cochlea as well as temporal coding of F0. These strategies are simulated in the acoustic domain using information known about electrical stimulation of the ear and functions of the auditory system.

6.8 Pilot experimental study

The NH subject who participated in this experiment was a native Cantonese speaker from Hong Kong, who lived in Australia but now resides in the USA and speaks English as a second language. The subject has NH levels and does not have a CI.

The Cantonese phoneme /si/ was chosen for investigation which has six different meanings depending on the tone used. The accepted tones and tone numbers are shown in the table below.

Tone #	Tone Name	Meaning	Character	Phrase	Meaning
1	High Level/Falling	Poem	詩	一首詩	A poem
2	High Rising	History	史	歷史	Past history
3	Middle Rising	Test	試	考試	An exam
4	Low Rising	City	市	大城市	A large city
5	Low Falling	Time	時	幾時	What time?
6	Low Level	Event	事	大件事	A big event

Table 6.1 Cantonese tones, meanings, characters, phrases used for the syllable /si/.

A tape recorder was used to record the subject. The subject recorded the six tones alone and the six tones in contextualised phrases. These data were then transferred to a computer, stored as 'wav' files, and filtered to reduce any noise or hiss. The speaker was the same as the subject and there may have been some advantage in recognising the pronunciation of the words on an absolute level but this study was primarily concerned with the recognition rate under different signal processing conditions examining relative recognition rates and thus thought not to affect the outcome.

These six words and six phrases were then processed using two schemes. The first scheme used a filterbank ("**FB**") strategy that filtered the sound data into sixteen channels using an FFT. The six channels with the largest energy were chosen and the audio output calculated by combining the appropriately weighted, filtered impulse responses from the chosen filterbank. This was calculated with a window period of 4ms (250Hz) similar to some current CI processing schemes. This has the effect of reducing the frequency resolution to 250Hz. The second scheme was the same except that the lowest filter in the filterbank was discarded and replaced with the extracted F0 frequency instead of the weighted filtered impulse response as was done in the first scheme. This scheme is referred to as (“**FB+F0**”). These data were then converted back into audio files.

The audio files were then presented to the subject for identification, in 6 groups. These groups were: ‘Words using raw audio’, ‘Words using FB’, ‘Words using FB+FO’ ‘Phrases using raw audio’ ‘Phrases using FB’ and ‘Phrases using FB+F0’.

In each of the groups the subject listened to the six words or six phrases, a total of 3 times each, presented in a random order. For each audio presentation the subject had to identify the correct word or phrase from a closed group of the six words or phrases.

6.8.1 Results

Spectrograms of the six tonal versions of /si/ are shown in Figure 6-1. It is clear that the broad spectral features of these six syllables are almost identical.

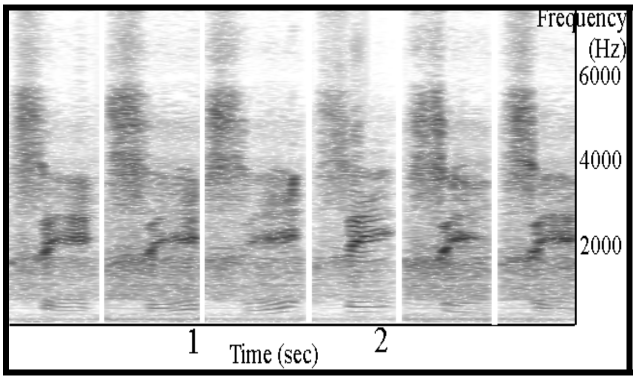


Figure 6-1 Spectrogram of the six different tonal syllables /si/.

A more detailed spectrogram of the low frequencies was used with the six tonal syllables and the plots of changes in voiced frequency are shown in Figure 6-2.

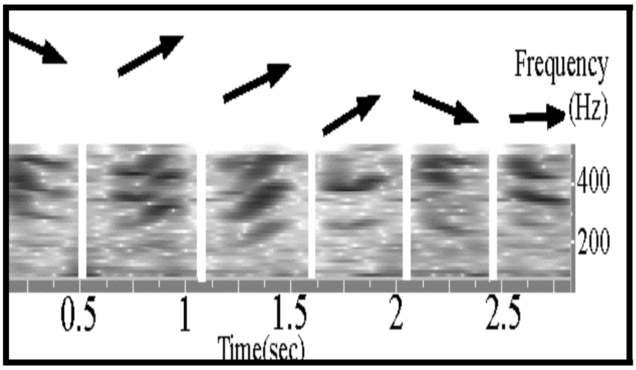


Figure 6-2 Detailed spectrogram of the six words used in the tests. Bar lines indicate the end of each word. Arrows indicate the direction of the change in pitch.

Results of word and phrase identification are shown in Figure 6-3.

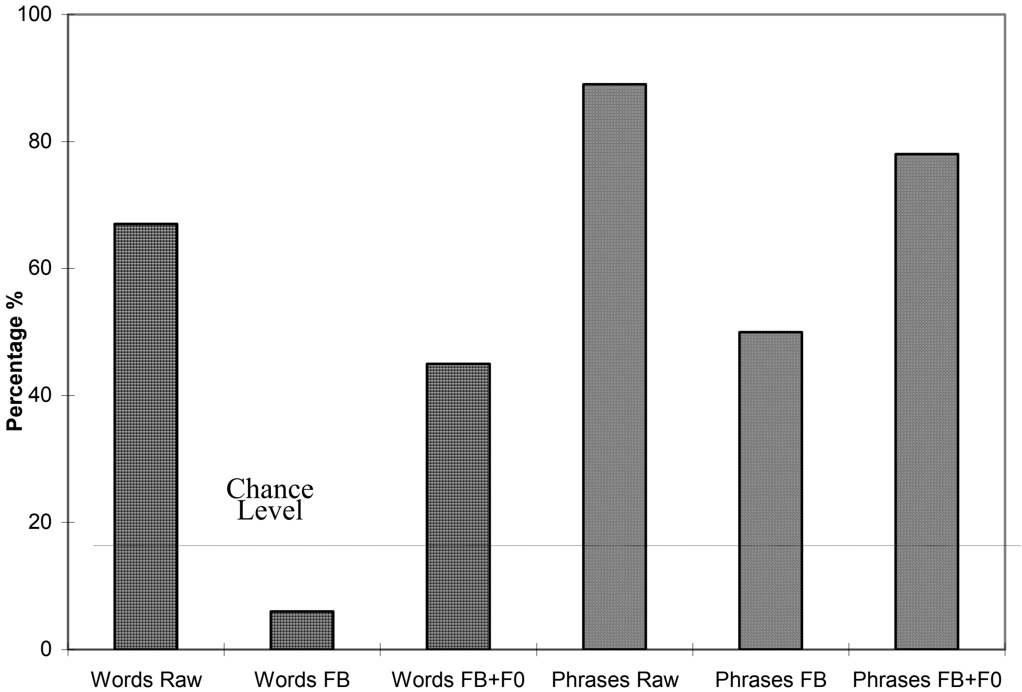


Figure 6-3 Word and phrase test results

The results clearly show that identification of these isolated unprocessed tonal syllables is reasonably difficult, with the mean recognition of raw audio being 67% (12/18). The identification of the tonal syllables drops to 6% (1/18) when processed by the FB strategy but improved to 45% (8/18) when processed by the FB+F0 strategy.

When the subject was presented with the tonal syllable in a short contextualised phrase, the mean recognition of raw audio was 89% (16/18). When the FB strategy was applied, the identification rate drops to 50% (9/18). When the FB+F0 strategy was applied, the identification rate improved to 78% correct (14/18).

6.8.2 Discussion of pilot study

The subject scored 67% in a closed set task using unprocessed audio files of the subject's previously recorded voice. This is well above chance of 17% and shows that the subject was using the other features of the syllable such as the pitch, register, voice quality, amplitude and duration.

When the FB strategy was used with the isolated words that did not deliver fine resolution of F0, the subject scored one correct out of eighteen (6%). This is well below chance and the subject communicated they were guessing as each word sounded very similar. The transmission of broad frequency information and transients appears to be adequate for the identification of the syllable but the loss of the tonal information by using the FB strategy is demonstrated by the below chance score in identifying which tonal variation of the syllable. When this strategy was changed to include finer resolution F0 details, as may be done by F0 rate presentation to CI users, the subject's score improved dramatically from 6% to 45%.

The identification of the original unprocessed phrases was 89%. This demonstrates how the contextual cues improve the recognition of the tonal syllable. When the phrase was processed by the FB strategy the identification dropped to 50%, which was much improved on the 6% scored when using isolated words alone. The mean identification improved to 78% when the phrase was processed by FB+F0. All the phrase scores were higher than the single word scores indicating the importance of context. There are two possible contextual cues that may be improving the phrase scores in this part of the experiment. The first contextual cue is the linguistic context that also occurs in non-tonal language. Secondly, there may be a contextual cue in a reference pitch provided by the phrase.

6.8.3 Conclusions of the pilot study

Audio simulations of the electrically stimulated ear can never fully represent what the CI user might hear. The actual separation of rate and place of stimulation information that a CI user may receive is virtually impossible to reproduce in the acoustical domain. These simulations may however, provide some idea of what parameters may be important in a strategy and are worth pursuing with a full study using CIRs.

The results demonstrate the importance of F0, namely, that without F0 extraction the identification of phonetically identical words with different tones is similar to chance. CI users speaking tonal languages may score better on contextual sentences than isolated words. Contextual sentences allow a subject to use the surrounding words to assist in determining the meaning of the word. This ambiguity in word discrimination may reduce speech rate understanding and subtract from the already reduced information that CI users receive. Strategies that carry information that allows the recognition of differences may improve the overall comprehension of tonal languages.

6.9 Further experiments with simulations

The next stage of the study used two native Cantonese speakers and two native Thai speakers. The aim of the experiment was to gain insight into the speech processing schemes that would be required for these different tonal languages, and to increase subject numbers; since data from just one subject in a pilot study is clearly insufficient to make decisions on the changes to speech processing strategy required. Further, the creation of a processing scheme that worked for different tonal languages would be better than just focusing on one tonal language that might limit the application and distribution of a new speech-processing scheme.

The simulation scheme was similar to that in the pilot experiment except that the number of channels was increased to 22 (previously 16) and the number of maxima was increased to 8 (previously 6) to keep in line with current CI processing schemes. The contextualised phrases used in the pilot study were not used in this next stage for two reasons. Firstly, they did not provide any more information than the isolated words provided and secondly, that the duration of presentation and number of words in each phrase was difficult to keep constant and thus added an unwanted confounding variable.

For the Cantonese speakers the syllable /si/ as well as /fu/ were used. Each syllable has six different meanings depending on the tone used as shown in below (/si/ had been used in the pilot experiment but in this study a different recording was used).

Tone #	Tone Name	Meaning	Character
1	High Level/Falling	Husband	夫
2	High Rising	Tiger	虎
3	Middle Level/ Rising	Trousers	褲
4	Low Falling	To Hold/Support	扶
5	Low Rising	Woman	婦
6	Low Level	Father	父

Table 6.2 Tones, meanings and characters for the Cantonese word /fu/

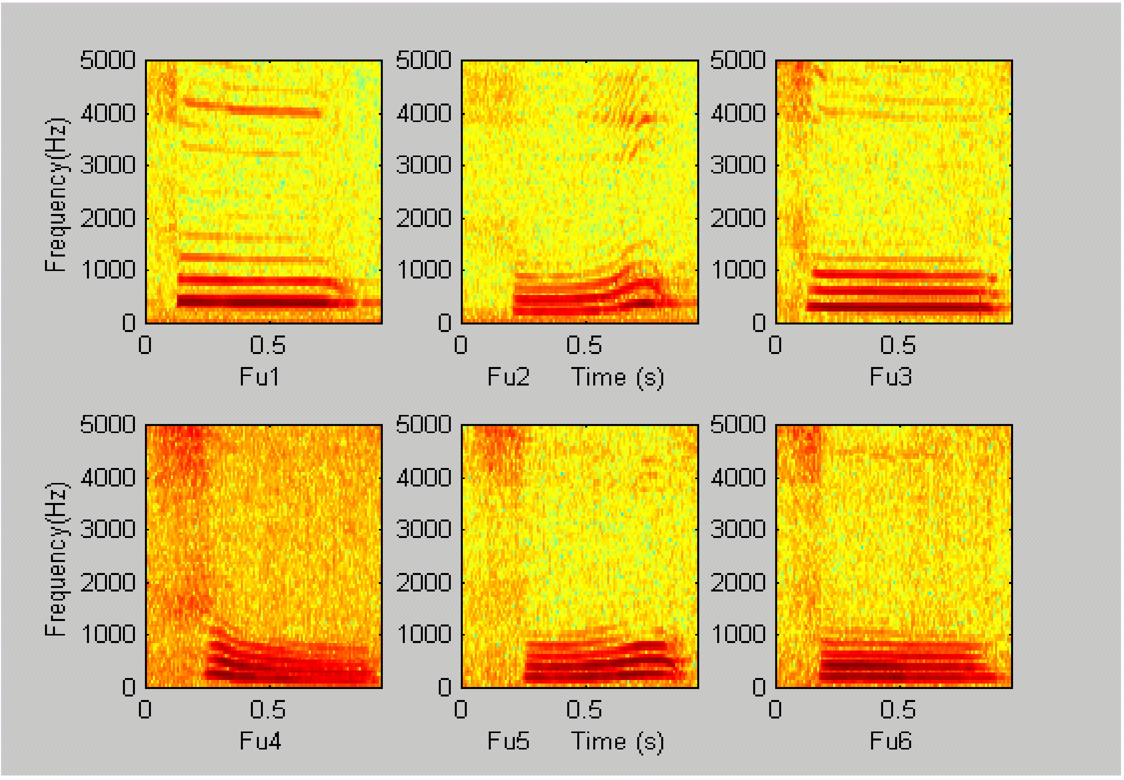


Figure 6-4 Spectrograms of the Cantonese syllable /fu/

Tone #	Tone Name	Meaning	Character
1	High Level/Falling	poem	詩
2	High Rising	history	史
3	Middle Level/ Rising	test/try	試
4	Low Falling	time/hour	時
5	Low Rising	city	市
6	Low Level	event	事

Table 6.3 Tones, meanings and characters for the phrase using the Cantonese word /si/

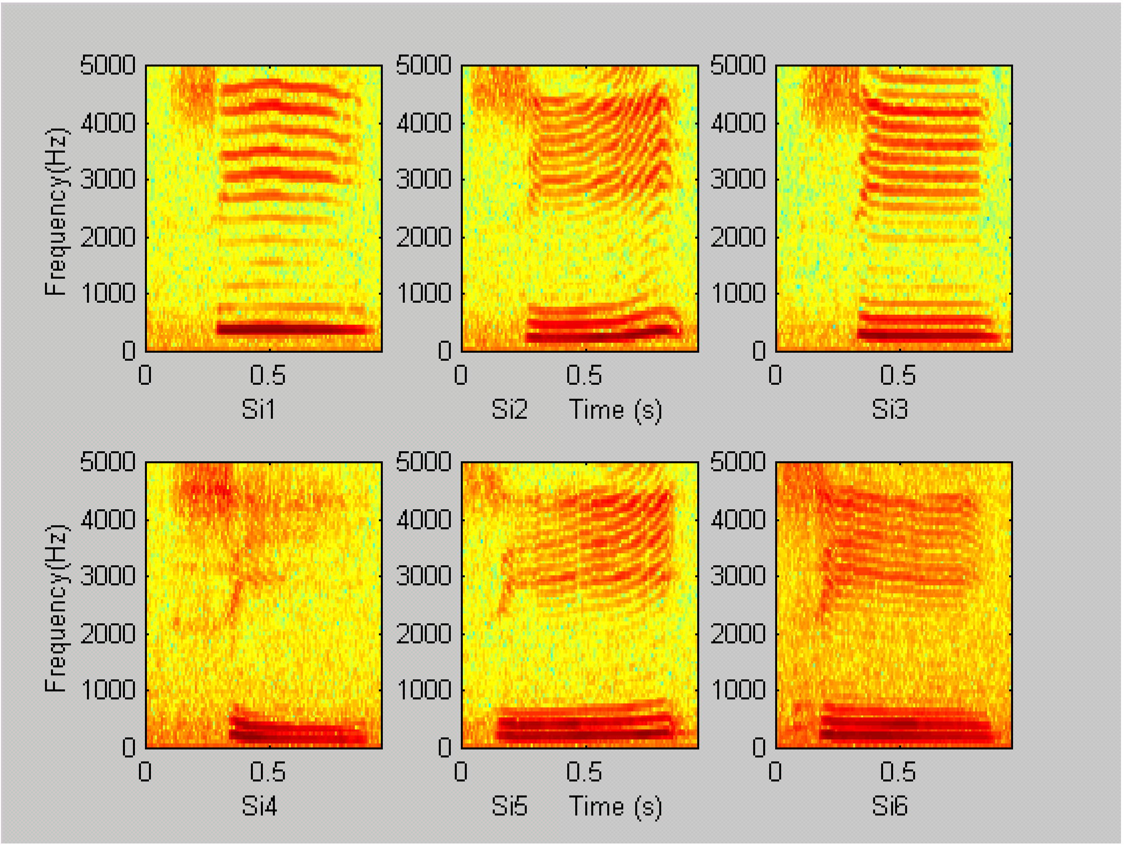


Figure 6-5 Spectrograms of the Cantonese syllable /si/

For the Thai speakers the words /aa/ and /khaa/ were used. The Thai language has very few occurrences where all tones can be used with same syllable phoneme. To overcome this, the tone names were used instead of word meaning (as some of the tone variation would not have a meaning). The subjects were given time to practice so that they could familiarise themselves with the task.

Tone #	Tone Name	/khaa/	/aa/
1	mid (monotone)	ขา	อา
2	low	ข่า	อ่า
3	falling	ข้า	อ้า
4	high	ข๊า	อ๊า
5	rising	ข๋า	อ๋า

Table 6.4 Tones and characters for the Thai syllables /aa/ and /khaa/

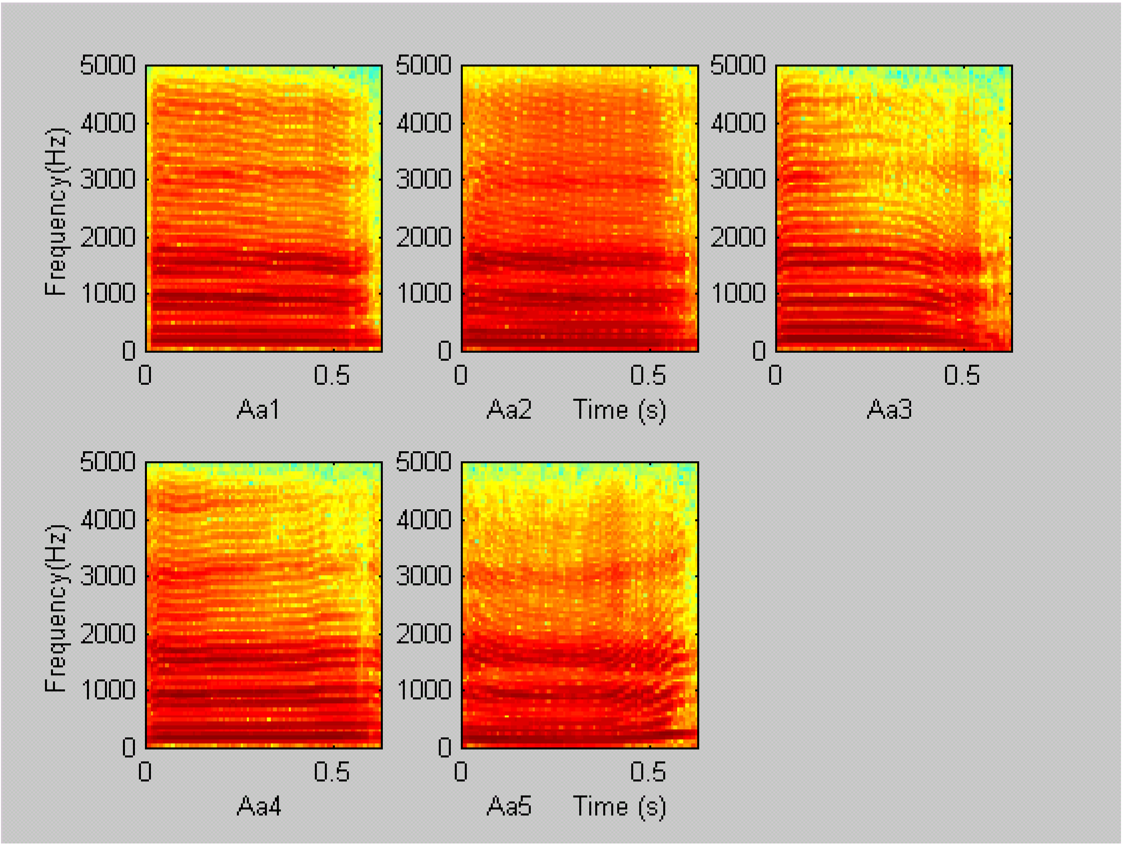


Figure 6-6 Spectrograms of the Thai syllable /aa/

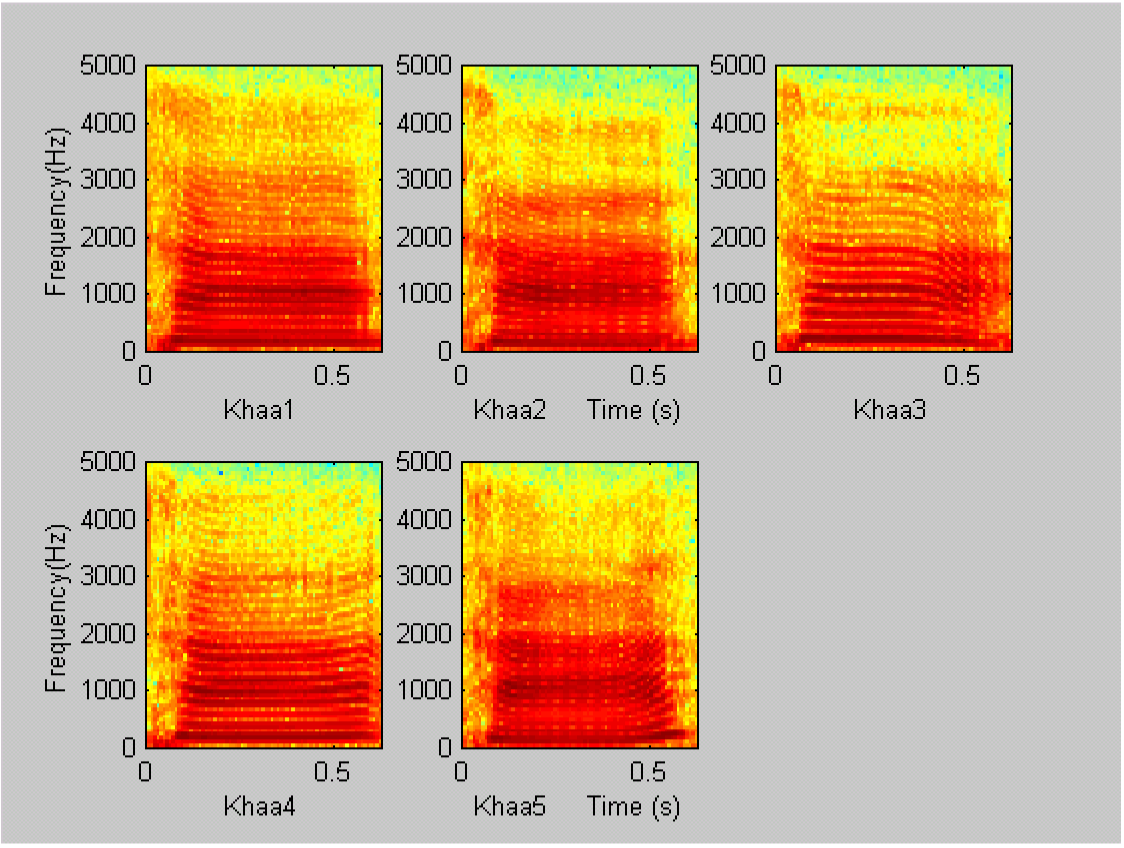


Figure 6-7 Spectrograms of the Thai syllable /khaa/

The Cantonese and Thai speaking subjects were presented with either the six Cantonese or five Thai words, respectively, twice in a random order. The subject was prompted to identify which tone was just heard, before presenting the next word. This was repeated for each of the processing schemes: Raw, FB (Filterbank strategy) and FB+F0 (Filterbank with added F0).

6.9.1 Results of Further Experiments with Simulations

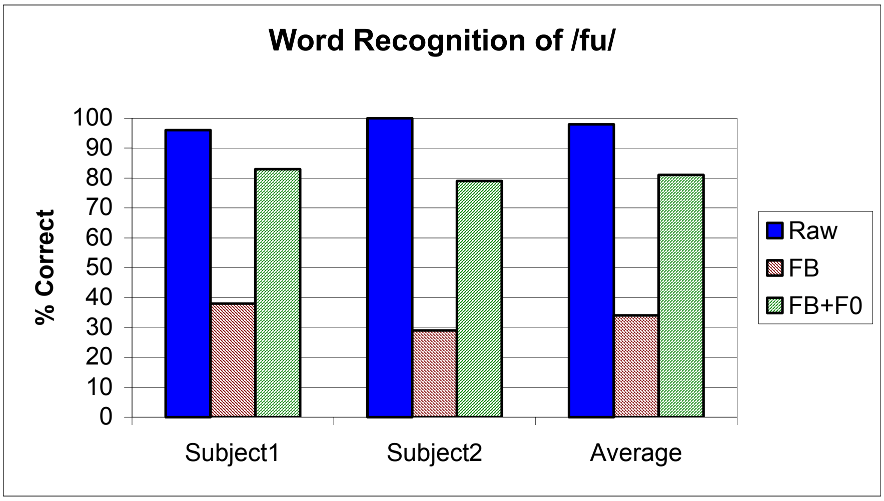


Figure 6-8 Results of Cantonese tone recognition test

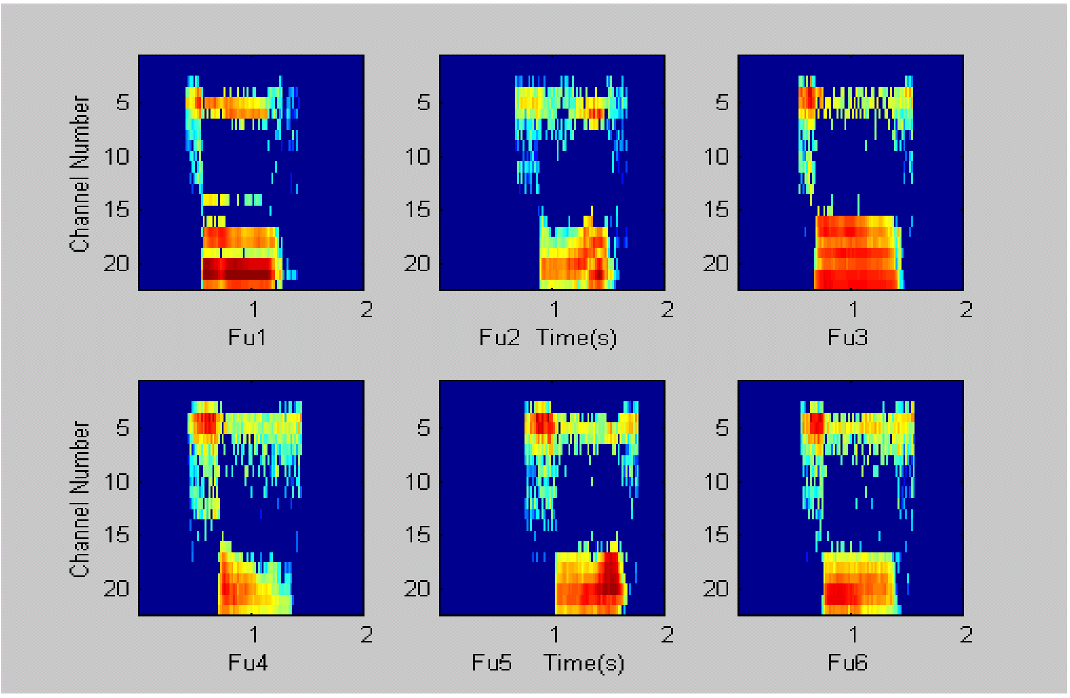


Figure 6-9 Electrodogram using ACE speech processing strategy of tonal variation of the word /fu/

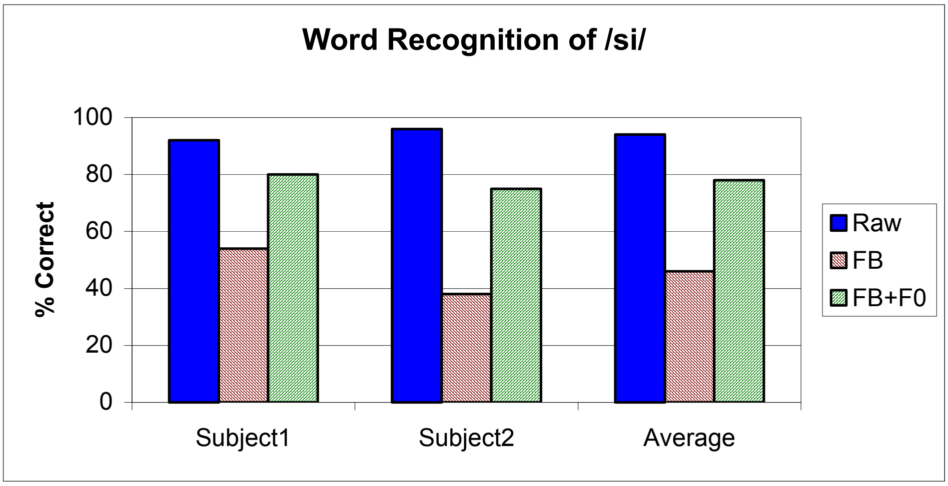


Figure 6-10 Results of Cantonese tone recognition test

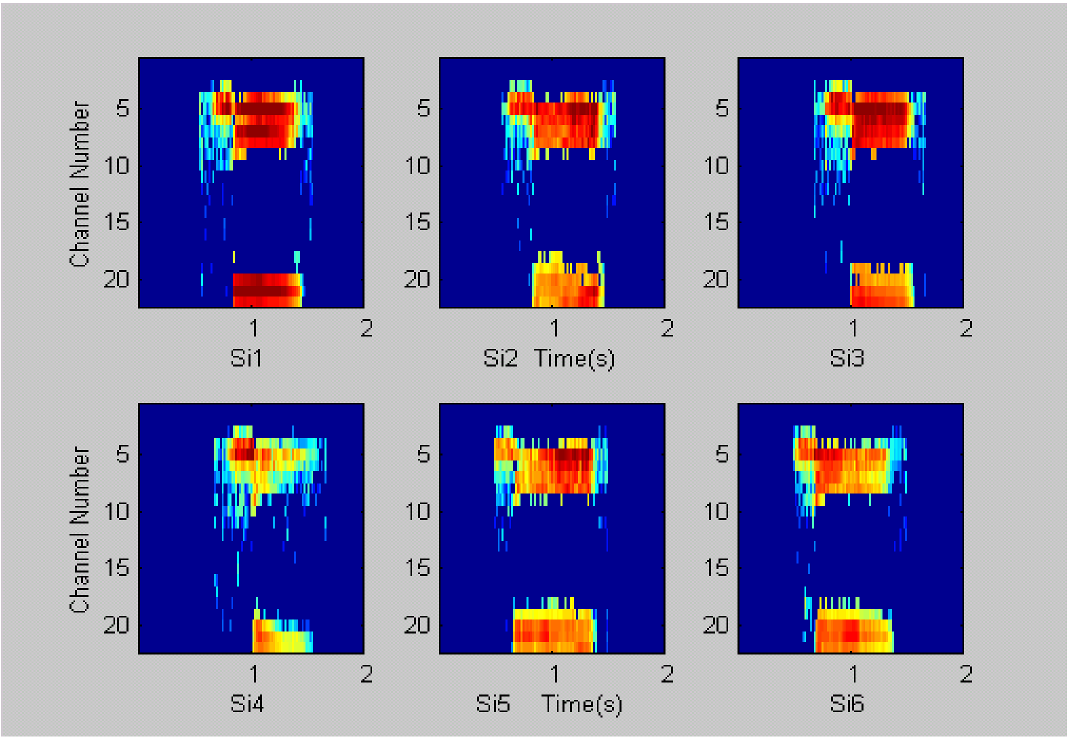


Figure 6-11 Electrodogram using ACE speech processing strategy of tonal variation of the word /si/

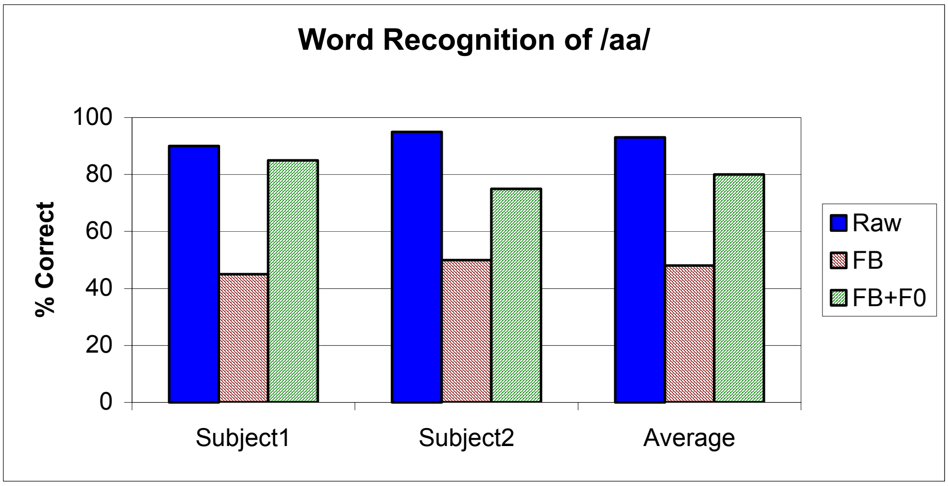


Figure 6-12 Results of Thai tone recognition test

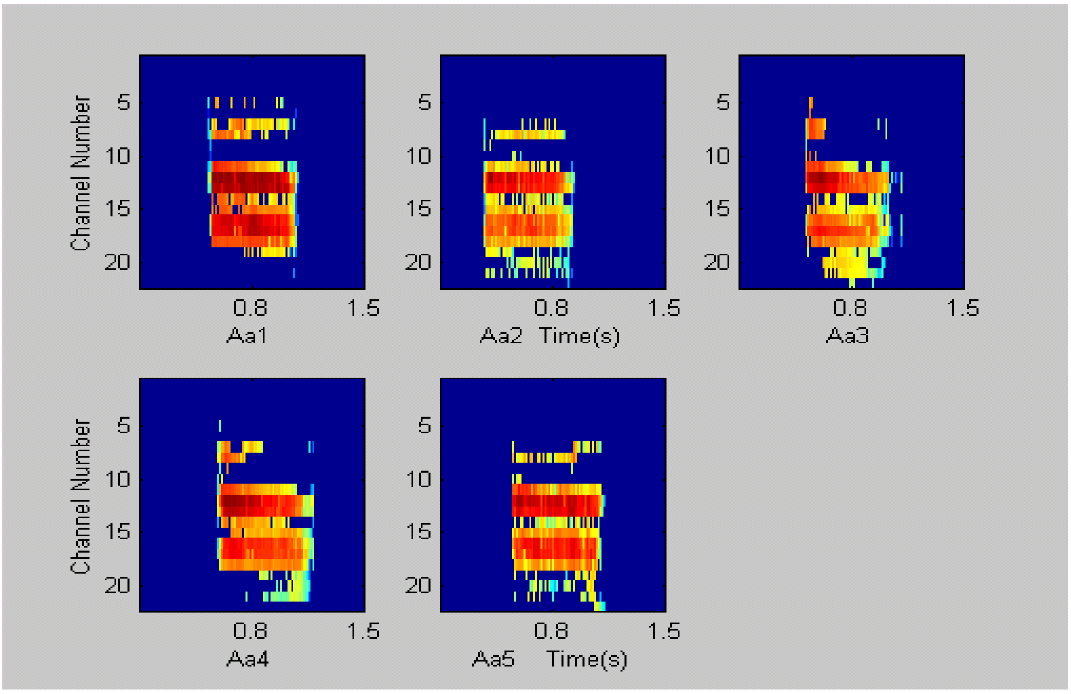


Figure 6-13 Electrodogram using ACE speech processing strategy of tonal variation of the Thai word /aa/

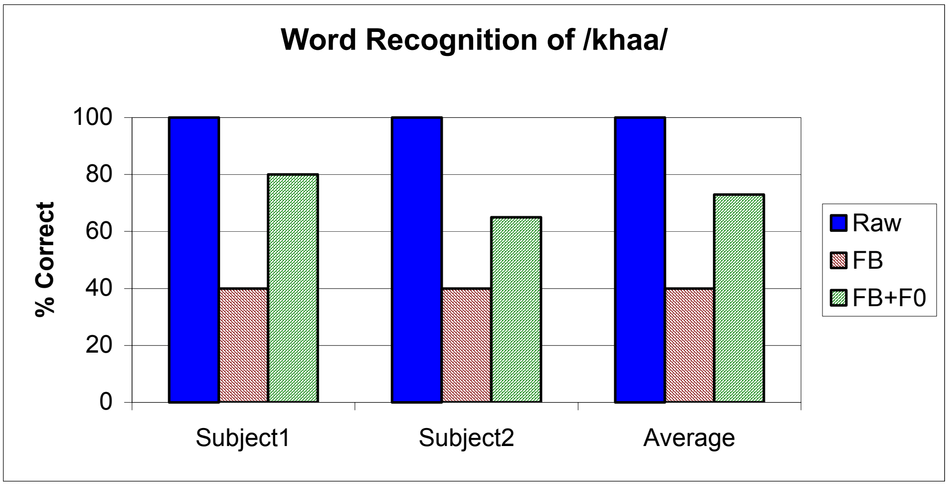


Figure 6-14 Results of Thai tone recognition test

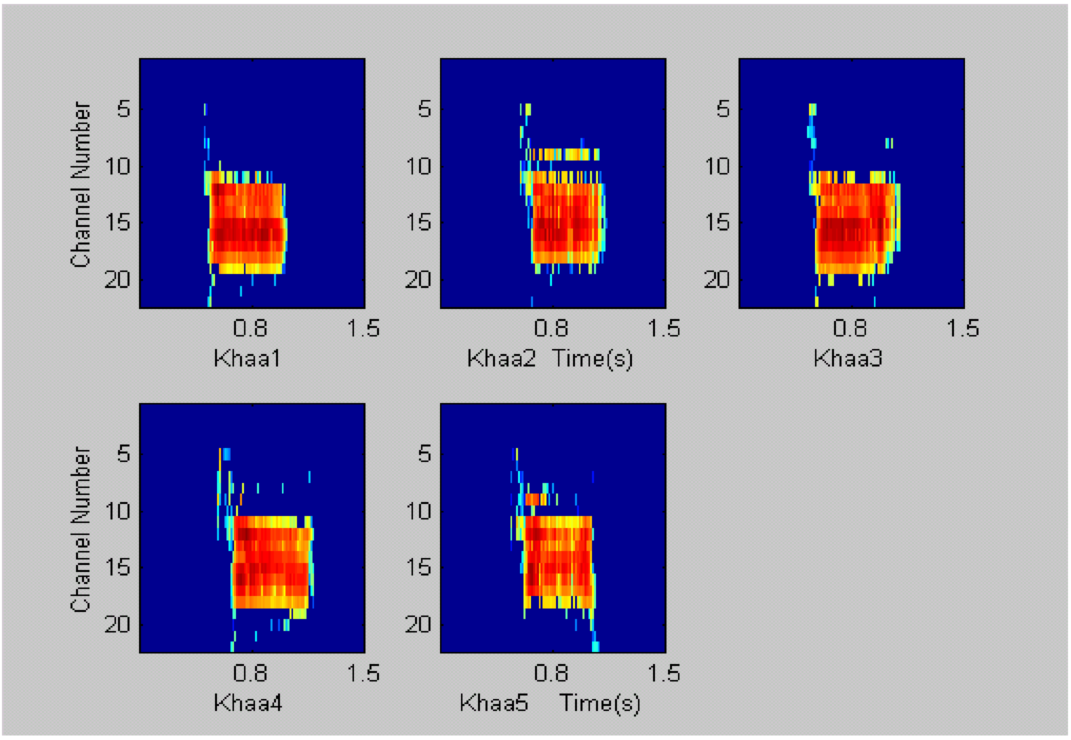


Figure 6-15 Electrodogram using ACE speech processing strategy of tonal variation of the word /khaa/

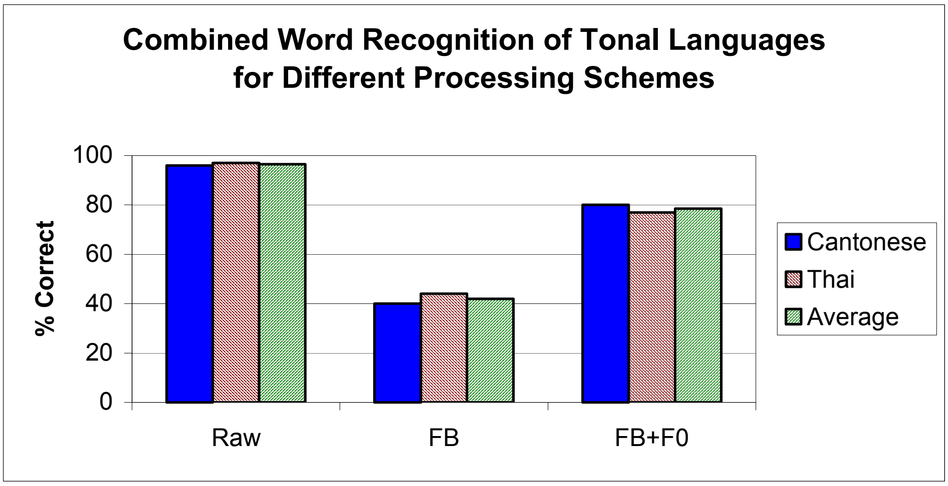


Figure 6-16 Combined results of Thai and Cantonese tonal tests

For the Cantonese phonemes /si/ and /fu/, the most common errors when using the FB processing scheme were that the level tones (high, middle and low) were confused and that the rising tone (high and low) were confused. The rising tones were identified as being different from the level tones with a greater accuracy than were the tones within each group (level or rising).

For the Thai phonemes /aa/ and /khaa/, most recognition errors occurred amongst the three level tones but all three level tones were discriminated well from the rising tone. The falling tone was discriminated poorly compared with both the rising tone and the level tones.

6.10 Conclusion

It is clear from the results in this chapter that the normal filterbank type speech processing strategy, with arbitrary stimulation rate, may not be providing the necessary cues to determine pitch differences and direction of pitch change. These simulations, although not being a true representation of what a CIR would hear, give some idea of how speech-processing strategies almost eliminate pitch, which is almost unnecessary for speech perception in non-tonal languages such as English. The pilot experiment showed the degradation of perceived pitch with a Cantonese speaker and the further experiments confirmed this trend not only with more Cantonese speakers but speakers of another tonal language Thai. The reintroduction of the fundamental frequency appears to be one method of providing simple pitch cues to improve word and tone discrimination. It is suggested that a speech processing strategy to be used by speakers of tonal languages should include some improvement. The addition of the fundamental frequency is likely to improve the perception that is necessary for these tonal languages.

Chapter 7
Pitch discrimination of Pure and Complex Tones: Differences in the Limens

7.1 Introduction

Sek and Moore (1995) investigated with NH listeners the difference between the ability to detect differences in pitch, compared to the ability to detect the direction of pitch. To differentiate these they termed them: difference limens for change (DLC) and difference limens for frequency (DLF) respectively. Their results showed that the DLF and DLC, while being dependent on frequency, were almost identical for NH listeners. In most investigations with CIRs the DLC is often quoted at the pitch discrimination level. It has been observed that while CIRs perform well on speech discrimination tasks they often confuse the concept of pitch with loudness. Pitch reversals are also commonly perceived even when the stimulus, a musical scale, has not reversed in directions (e.g. only ascending or descending). It is hypothesised that the DLC and DLF for CIRs might be quite different unlike NH subjects due to differences found in speech and music perception found with CIRs. DLC may be more akin to speech comprehension, as the CIR only needs to determine if signal has changed and be able to reliably recognise this change. DLF may be more useful as a measure of music perception due to the required ability to indicate the direction of change in pitch.

New psychophysics software was written to measure these difference limens:

i. The Just Noticeable Difference (JND) test, which measures difference limens for change.

ii. Pitch Discrimination Test (PDT), which measures the difference limen for frequency.

All tests used in this chapter presented sound acoustically. This allows the recipient's processor to process the sound and is a measure not only of the CIR's ability but also of their current processor, microphone and implant.

7.2 Aims

What is the difference between levels of difference limen for change and difference limens for frequency using complex and simple stimuli?

It has been observed that CIRs are often unable to indicate the direction of pitch. CIRs who are able to distinguish if two tones are different may still have difficulty in indicating the direction of pitch change. This remains to be tested. This may be a possible cause of being unable to recognise once familiar music, while the absolute discrimination level may be responsible for being able to discriminate new music.

The dependence of pitch discrimination ability on timbre of the stimulus has not been tested with CIRs. It is hypothesised that timbre may have some effect on the pitch discrimination levels due to the differences in the number of electrodes that are stimulated and the effect of a peak picking algorithm. Often the fundamental frequency is not one of the largest peaks, which may be due to the type of stimulus (for example brass instruments often have harmonics which are stronger than the fundamental), to reproduction on a smaller loudspeaker, or to the microphone or pre-emphasis of the speech processor. This can lead to the fundamental frequency not being selected in a peak picking algorithm and therefore not being presented to the CIR. This may have no effect when the CIR judges whether two stimuli are the same or different but may increase the difficulty when required to judge which stimulus is higher or lower in pitch. NH listeners do not always require the fundamental to determine the pitch of the stimulus. The resolution available to the NH listener at the locations in the ear where the harmonics stimulate (as well as those introduced distortion products) often provide ample information to form a pitch estimate.

To test this hypothesis two different types of stimuli were used. They are:

i. A Sampled Piano

ii. Pure Tone

What is the difference in pitch discrimination levels achieved between CIRs and normal hearing participants?

It is to be expected that the frequency difference limens will be much greater for the CIR than for the NH population. This may be due to the reduced resolution available to the CIR caused by the reduced nerve survival and discrete nature of electrodes. The electrode array also does not span the full length of the cochlea. A redistribution of frequencies is done to compress frequencies normally spanning the 35mm length of the cochlea into 22 electrode sites covering 25mm. This frequency distribution is contained in the ‘map’ that can be individual for each recipient.

What is the dependence of the difference limen on frequency?

It is known that the frequency difference limen is dependent on frequency for NH listeners. It is of interest examine the dependence on frequency of the frequency difference limens for CIRs.

What is the effect of speech processing strategy type upon pitch discrimination?

One of the subjects available for this experiment uses both the ACE and SPEAK speech-processing strategy. It is of interest to investigate whether the type of speech processing strategy used may have an effect on the ability to discriminate pitch. Speech-processing strategies that have been developed to improve speech perception may then be customised to individual CIR pitch discrimination requirements.

7.3 Psychophysics Tests

7.3.1 The Just Noticeable Difference (JND) test

The JND test presents a series of four tones to the participant and asks them to identify the tone that is different in pitch, a Four Alternate Forced Choice task (4AFC). This test yields one of the standard measures of pitch discrimination. It does not require the participant to indicate the direction of difference of the pitch i.e. higher or lower, but just which tone is different. This test is shown in Figure 7-1.

7.3.2 Pitch Discrimination Test (PDT) of isolated frequencies - Difference Limen for Frequency (DLF) test

The difference limen for frequency test requires the participant to identify the direction of pitch of two presented tones by indicating if the second tone was higher or lower in pitch than the first. This test is shown in Figure 7-2.

7.3.3 Preliminary investigation

Before performing the main set of experiments, a preliminary investigation needed to be undertaken to determine the effect of one of the parameters used in the main experiments. In most pitch discrimination studies, the loudness is usually randomised, or roved, to reduce any cues for pitch discrimination that may be made available by differences in loudness. The researcher is posed with the problem of needing to randomise the loudness to reduce the available cues that may aid in the discrimination of pitch but simultaneously wishing to separate two experimental variables that may both affect pitch discrimination. What is the optimum amount of variation in the loudness to reduce any available cues to aid discrimination without degrading the ability to detect changes in pitch?

To investigate this effect, a NH participant was tested by varying the amount of loudness randomisation and monitoring the JND level.

7.3.4 Stimuli

The pure tone and piano stimulus were chosen for this experiment. The pure tone is of interest as it contains, by definition, a single frequency. Any differences detected by the recipient will then be due to changes in this single frequency and will be reflected in a speech processing strategy by changes in electrode stimulation patterns. The piano stimulus is of interest as it is a common musical instrument and in contrast contains multiple harmonics. The peak picking techniques of current speech processing algorithm will select harmonics with the highest maxima, which may not include the fundamental. With NH, this pitch can be derived by various proposed methods including harmonic spacing, distortions of the basilar membrane or non-linearities in the neural pathways. The CIR with reduced pitch resolution ability may not be able to utilise these mechanisms and thus be unable to determine with any accuracy the pitch of the stimuli.

The tests used in these experiments are automated and self administered by the participant. The author wrote testing software using Microsoft's Visual Basic programming language. The testing software controls a Kawai synthesiser General MIDI (GM) module "GMega" and the stimulus was delivered acoustically via a Yamaha YST-40 speaker. The tests were performed at frequencies in the vicinity of 200, 400, 800, 1600 and 3200 Hz. The frequency of the standard MIDI note closest to these frequencies (i.e. G3-G7) is used as the standard tone to simplify the programming for the computer controlled MIDI module that is used.
These frequencies are shown in Table 7.1.

Standard Note Name	Midi No	Frequency (Hz)	MIDI Note Name
G3	55	196	G5
G4	67	392	G6
G5	79	784	G7
G6	91	1568	G8
G7	103	3136	G9

Table 7.1 Frequencies and MIDI notes used.

For the JND test that measures DLC the test uses a 4AFC adaptive procedure similar to that of Gfeller et al. (2001 Cochlear Implant International, in press) but which used pure tones only. For the PDT isolated frequency test a 2AFC adaptive procedure is used. Both tests were conducted:

i. In a soundproof booth to limit external noise and to reduce reverberations and standing waves that may occur and influence the discrimination ability.
ii. Using randomised loudness levels of the tones presented over a range of 4dB to minimise loudness cues that may aid discrimination ability.
iii. Using a comfortable loudness for the test participant, this is typically 70-75dB SPL.

Before beginning the test the participant was encouraged to use the “practise” button to hear example stimuli with feedback in order to be accustomed to the task. This practice was found to reduce any initial inaccuracy due to the participant being unfamiliar with the exercise. The practice period also gave the participant an opportunity to adjust the volume to a comfortable level.

The author advised the participant to listen for differences in pitch and not loudness. After the participant made their choice by clicking on the appropriate button on the screen, automatic feedback was given to the participant indicating about whether they were right or wrong by way of a graphic tick or cross appearing as can be seen in Figure 7-1 and Figure 7-2.

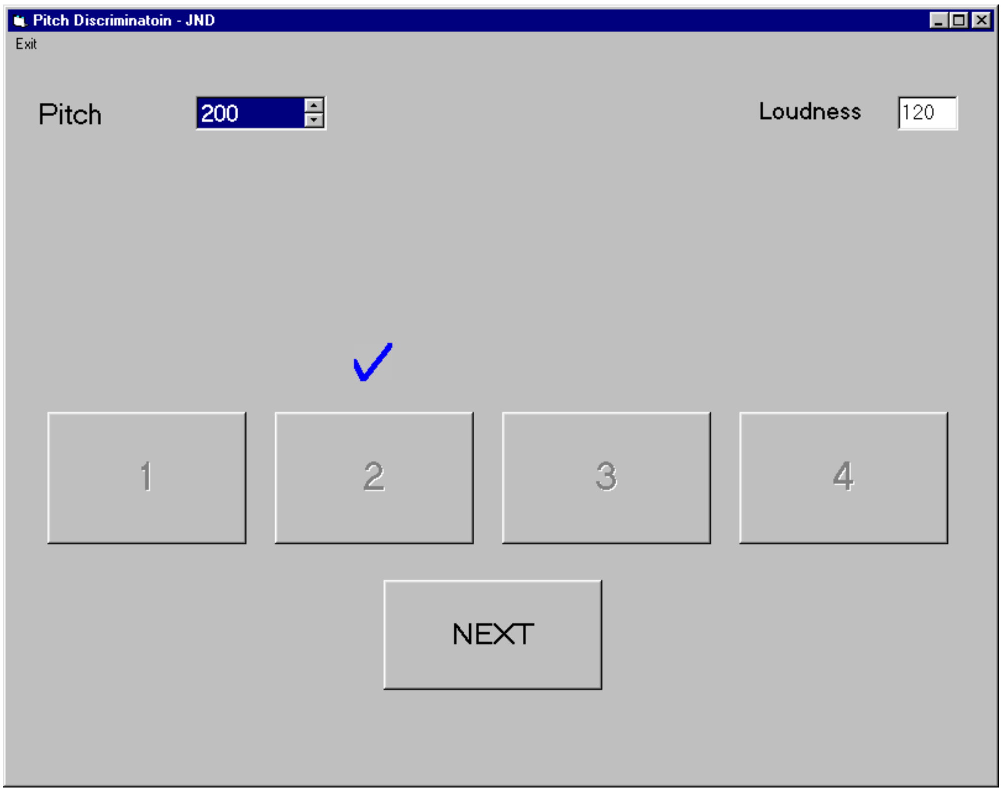


Figure 7-1 Screen shot from the JND testing software.

The pitch and loudness values were visible to the subject but were only used to select the base frequency and amount of loudness roving, which was not explained to the subject. These values did not change during the experiment and thus did not provide the subject with any information, which may aid their selections. It is quite normal for CIR participants to lose track of the characteristic they are trying to discriminate; they often begin to look for changes in loudness level. This is to be expected, as participants more frequently perform 'mapping' exercises that only require attention to loudness. The feedback provided by the test aims to reinforce the request that the participant concentrate only on changes in pitch.

An adaptive procedure was used to reduce the time necessary to test each participant and also to concentrate testing around the point of interest on the psychometric curve. The adaptive procedure used was a 2 down, 1 up scheme described in Kollmeier et al. (1988) and discussed in Chapter 4. The results could also be compared with results previously published by this author in collaborative work Gfeller et al. (in press) that only used pure tones as test stimuli.

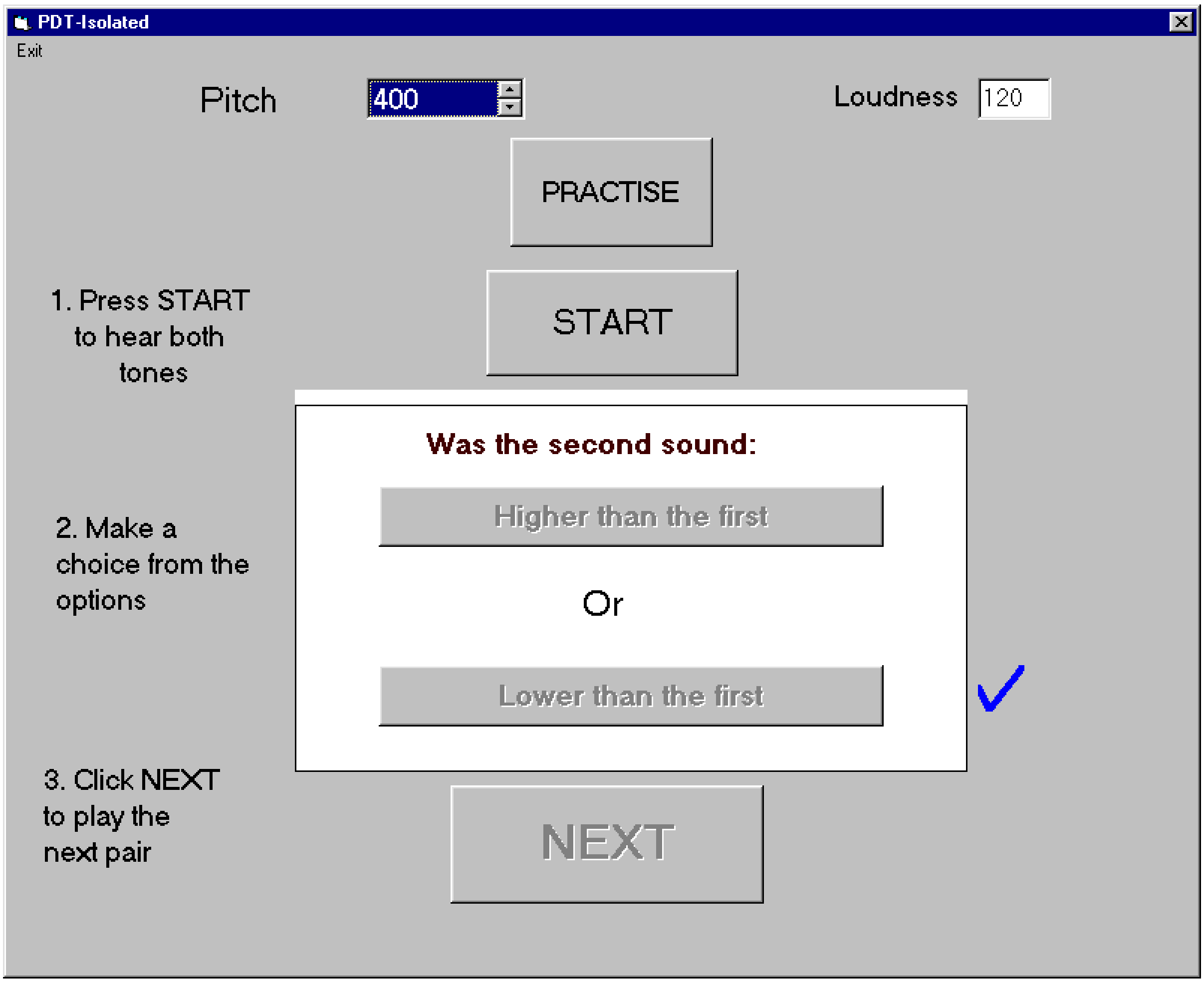


Figure 7-2 Screen shot of difference limen for frequency test

If the participant responded correctly to two presentations in a row, the step size was halved, but if the participant responded incorrectly the step size was doubled. Fourteen reversals were obtained from each run and the last ten reversals were averaged to give the frequency difference limen for that run. The fine pitch differences required in this test were accomplished using a feature in MIDI known as pitch bend. When a frequency is required that is not a whole note, which was quite common once the run begins to converge, a fine adjustment to the pitch was necessary. A pitch bend message is sent before each midi note that instructed the MIDI module to adjust the presentation pitch by a fixed amount. Each semitone could be further divided into 32 equal steps. The author developed a software algorithm to convert the desired frequency in Hertz to a MIDI note number and a MIDI pitch bend number. This produces a test system that has a pitch resolution of 0.19 semitones or 19 cents. The Kawai synthesiser has the standard "General MIDI" (GM) bank of instruments. Of these instruments available, the piano (acoustic grand piano) is used as a test stimulus. The GM synthesiser also has the ability to create sounds from various base oscillators. The sine wave (pure tone) is one oscillator available and was chosen as it consists of only a single frequency in contrast to the piano. This sine wave was created and the amplitude envelope programmed for attack and decay times that were 10 percent of the total length of the note (10 percent attack and decay times are standard settings used to eliminate the click that would be associated with a 0% attack or decay time). Stimulus duration in psychophysical experiments typically range between 200ms and 500ms (Tong and Clark, 1985; Tong et al., 1983a; Collins et al., 1997). In pilot trials stimulus duration of 250ms and 500ms were trialled with no difference found. Each tone was therefore set to be 250 ms separated by 250 ms of silence, although the software allows the selection of any length of stimulus and inter stimulus gap allowing for future experiments.

Each participant completed three consecutive runs in each frequency range with a randomised order. The final data consisted of the average of the last two runs. The average difference limen was calculated by dividing the smallest discriminable frequency by the test frequency to yield $\Delta f/f$.

7.3.5 Participants

Three CIRs took part in this study. Two use the ACE speech processing strategy and the other uses both the Speak and the ACE speech processing strategies. A participant description can be found in Table 7.2. All participants were postlingually deafened. Two subjects had greater than nine months experience and one subject had just over one month experience with their present devices. Three NH adults also took part in this experiment for a comparison.

Participant	Age	Implant Type	Processing Strategy	External Processor
CIR6	41	N24	ACE	Sprint
CIR7	63	N24	ACE SPEAK	Sprint Esprit
CIR8	47	N24	ACE	Sprint

Table 7.2. Participant details

7.4 Results

7.4.1 Preliminary investigation

In order to observe the effect of level randomisation a NH subject was tested at a frequency of 800 Hz under two conditions:

i. In the first condition, the loudness was kept constant. The participant recorded a difference limen of 4.5Hz at 800, or 0.56%.

ii. In the second condition the loudness was varied by up to 10dB. The difference limen increased to 18.3Hz at 800Hz, or 2.3%, an increase of 400%.

The results of this experiment indicate that allowing the sound level to vary by up to 10dB resulted in a marked reduction in frequency discrimination ability. Where is the ideal level that reduces the cue of loudness in detecting changes in pitch but also does not degrade the ability to detect changes in pitch by varying the loudness too much? In published experiments where difference limens are found for loudness the presented pitch was not varied at all (Rhode, 1998; Fingst et al., 1983). In comparison, almost all published experiments suggest the loudness needs to be randomised to find the difference limen for pitch (Townshend et al., 1987). Superficially this might suggest that pitch does not influence loudness judgments but that loudness influences pitch judgments. The curves published by Fletcher and Munson show, however, that this is not so (Fletcher and Munson, 1933). Indeed, it is believed that both loudness and pitch are interrelated, and experiments, as early as those conducted by Stevens (1935) described their interrelation. In published experiments on JND for pitch, the loudness is usually varied by around 6-8dB (Henning, 1965; Gfeller et al., 2001). In more recent studies Fu et al. (2001 Asilomar Conference) performed electrode discrimination experiments where the loudness was roved by only 2dB.

In this experiment a loudness randomisation value of 4dB was chosen to use with CIRs as it is within the range used in the literature but is on the lower side to reduce effect of loudness, masking cues of valid differences in pitch.

7.4.2 Results for DLC and DLF

The difference limens for the change and frequency were found for the piano and sine wave stimuli and are shown in Figure 7-3. An alternate presentation of the results grouped by stimulus type is shown in Figure 7-4.

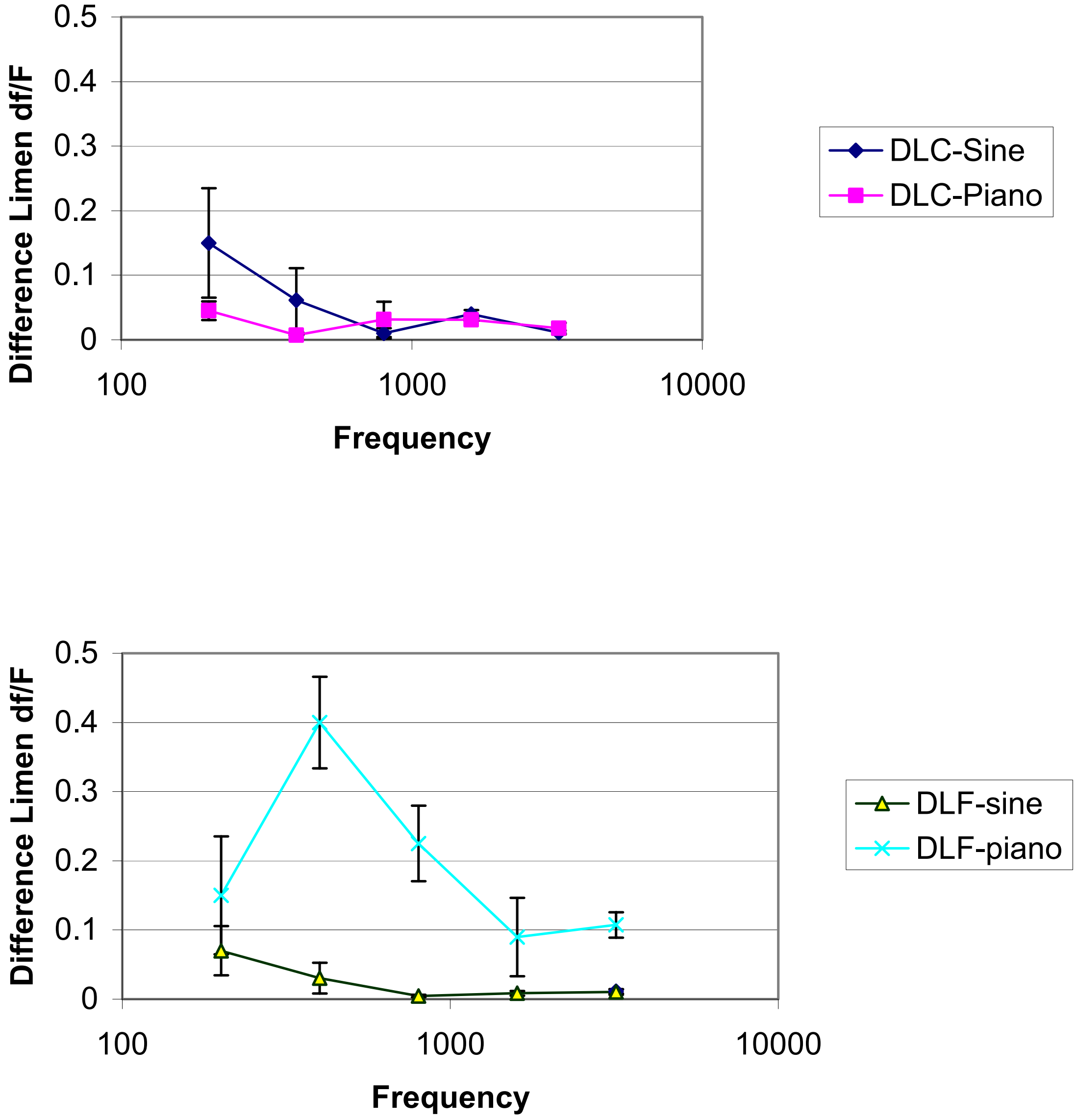
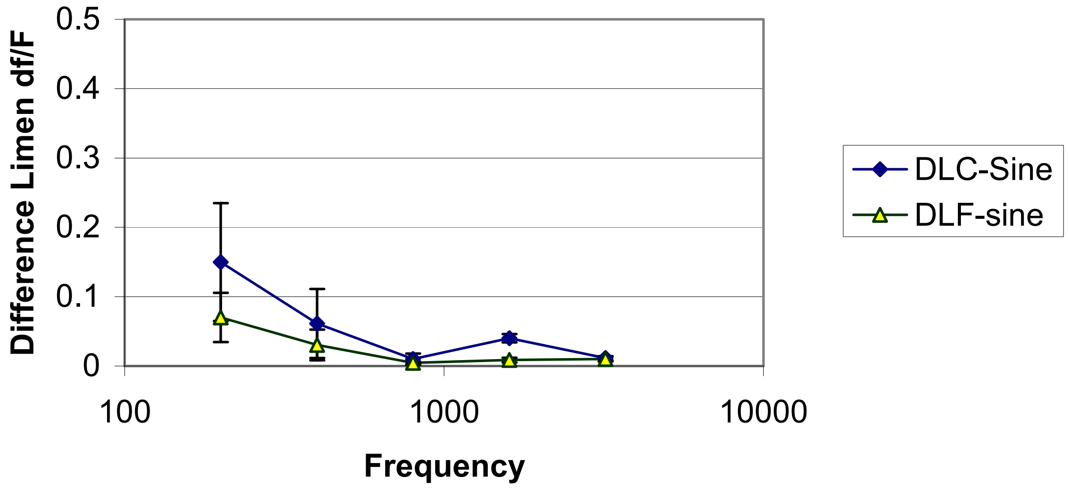


Figure 7-3 Difference limens for change and difference limens frequency for the pure tone sine and piano stimulus.



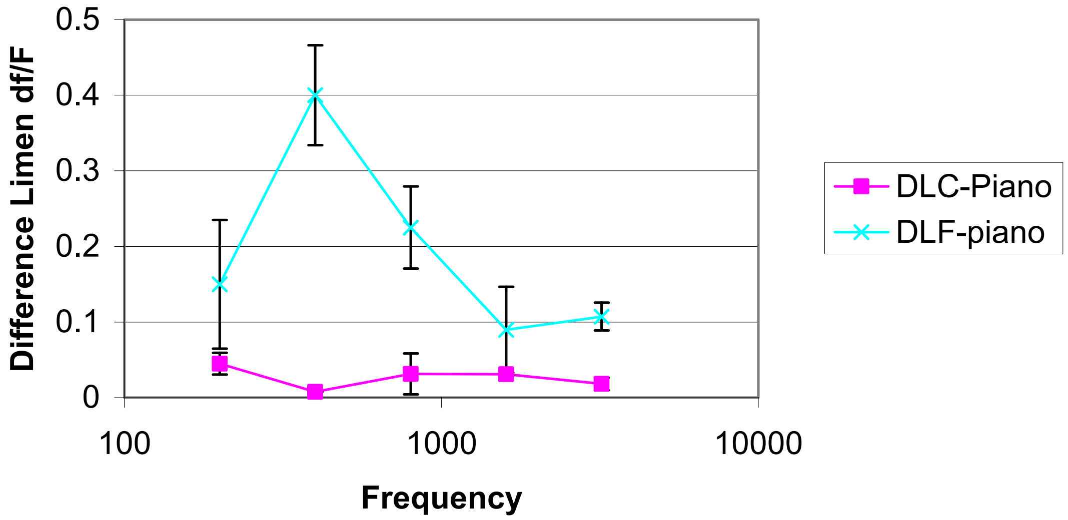


Figure 7-4 The same data as presented in Figure 7.3 but grouped according to stimulus. This is useful when comparing DLC and DLF on one graph at a time.

For sine waves, the difference limen for frequency and change are not significantly different (p=0.52). For the piano, however, difference limen for change is better (lower) than the difference limen for frequency (p=0.03). The difference limen for frequency for the sine wave is much lower than the piano (p=0.02). The higher difference limen for change, using the sine wave at low frequencies than the piano, is thought to be due to the lower number of electrodes/octave in the low frequency range. A low frequency sine wave will only stimulate up to possibly three low frequency electrodes where the filter/octave ratio is less. The piano stimulus, consisting of many harmonics, will stimulate many more electrodes. So while changes in the piano fundamental may barely change stimulation in the low frequency channels, the upper harmonics of the piano will lie in regions where the electrode/octave ratio is much higher. This will produce more changes in stimulation patterns for smaller changes in frequency than does the pure tone, which can be readily explained due to the harmonics stimulating more electrodes.

When the recipient is required to indicate the direction of pitch difference between the two tones the discrimination ability appears to be much better using the pure tone than the piano (p=0.02). This holds true over the frequency range tested. Changes in the pitch of the pure tones will be reflected in different amplitude and electrodes in a reasonably ordered fashion. As the frequency is increased, the electrode stimulation amplitude will increase until it finally reduces as the frequency leaves the respective filter. As the amplitude of this channel reduces the amplitude of the next (higher) frequency electrode will be increasing. However, there appears to be much more pitch ambiguity when the stimulus is changed to a piano.

In a practice session one subject reported that at 400Hz almost every interval sounded in the opposite direction and when the exercise was repeated the subject performed much better when they chose the direction the opposite of what they heard. This data was regarded as training and not included in the data presented here. This ambiguity may result from two possible sources:

i. The peak picking strategy does not always produce an increase in stimulation of electrode frequency for an increase in stimulus frequency. It can be seen from electrodograms shown in the discussion (this chapter), that there are situations where the resulting electrode stimulation pattern could be thought to be due to a change in pitch in the opposite direction.

ii. The current flow that is associated with a change in pitch may actually flow at various electrode positions via pathways that may stimulate nerves so that the pitch appears to be reversed. This would be thought to be individual for each subject based on their physiology, electrode insertion depths etc. The participant that experienced the pitch reversal at 400Hz for the piano, however, had very good discrimination ability for frequency when the stimulus was a pure tone.

The surprising result, and one, which may affect the way music-processing schemes are designed in the future, comes from comparing the DLC and DLF for the piano stimuli, shown in Figure 7-4. There is a much larger difference between the ability to detect differences in piano tone with different fundamental frequencies compared to the ability to detect the direction of the change in frequency. This result may explain why music may sound so bad for so many implant recipients. In a laboratory situation CIRs can detect changes in pitch of ~1 semitone. When hearing a complex tone they discriminate direction poorly; subjects report that they know the pitch has changed, but do not know how it has changed. In real music, discrimination is possibly worse still. When CIRs were asked what they were hearing, two clearly different explanations were offered:

i. "Sometimes, it is clear the instrument is the same it just appears that the pitch direction is the opposite than what was perceived".

ii. Secondly, and perhaps more interestingly, is that the pitch does not clearly change in pitch direction. CIRs have described this as "the same pitch played by a different instrument" or some have described it as "part of the note goes up and part of it goes down so I am just guessing at which way the pitch has changed'.

Most scientific experiments on CIRs use a pure tone and hence have found levels that, while much larger than those of NH listeners, are possibly enough to follow changes in simple melodies. The literature rarely reports studies where DLF are found. Of those that do, only the pure tone is used and the hence the results are similar to DLC, as was found here (Figure 7-4). However using this DLF test with a stimulus that subjects would encounter in everyday life like the piano, the discrimination abilities actually appear to make the directional changes in the melody ambiguous or incorrect so the melody appears to be unrecognisable. This result is considered to be the first discovery of this difference in levels achieved and no comparable test or results exist in the literature.

7.4.3 Results of differences due to strategy

The subject that uses both the SPEAK and ACE strategy was tested using the JND test measuring DLC. The results of the sine wave and piano stimulus have been combined to examine the effect of strategy and are shown in Figure 7-5.

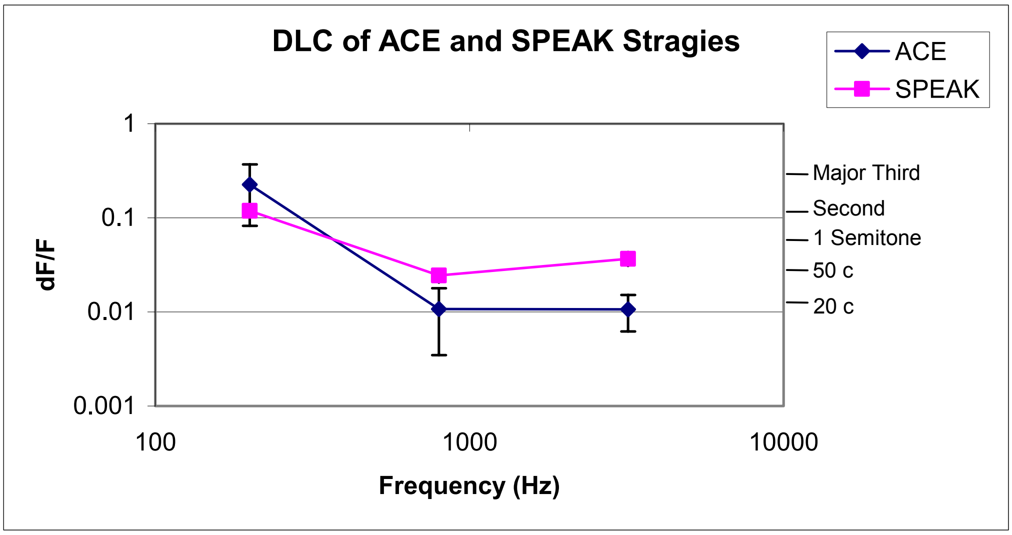


Figure 7-5 Frequency difference limens found using different speech processing strategies

It appears from the data that, for two higher frequency samples (800Hz and 3200Hz), the frequency difference limens are lower using the ACE strategy than using the SPEAK strategy. Superficially, ACE would seem to have an advantage for music, however, it should be noticed that frequency difference are large in the frequency range associated with music.

Comparing the results from this experiment with the results of Gfeller (2001) it appears the frequency difference limens at 200Hz are around average compared to the sixteen participants used in that study. For the other frequencies of 800Hz and 3200Hz it appears that the frequency difference limens for these participants from this study are equal to those of the best participant recorded in the literature.

7.4.4 Results of normal hearing listeners

The data collected from NH listeners and the CIRs are shown in Figure 7-6.

The large difference between the difference limens between these two groups, significant at the $p<0.05$ level, is not surprising but serves to illustrate the large difference in pitch discrimination ability between CIR and NH participants. It is no wonder that tasks such as melody recognition and music perception in general are performed poorly by CIRs who have reduced ability in an area of musical skill so fundamental, that is normally taken for granted.

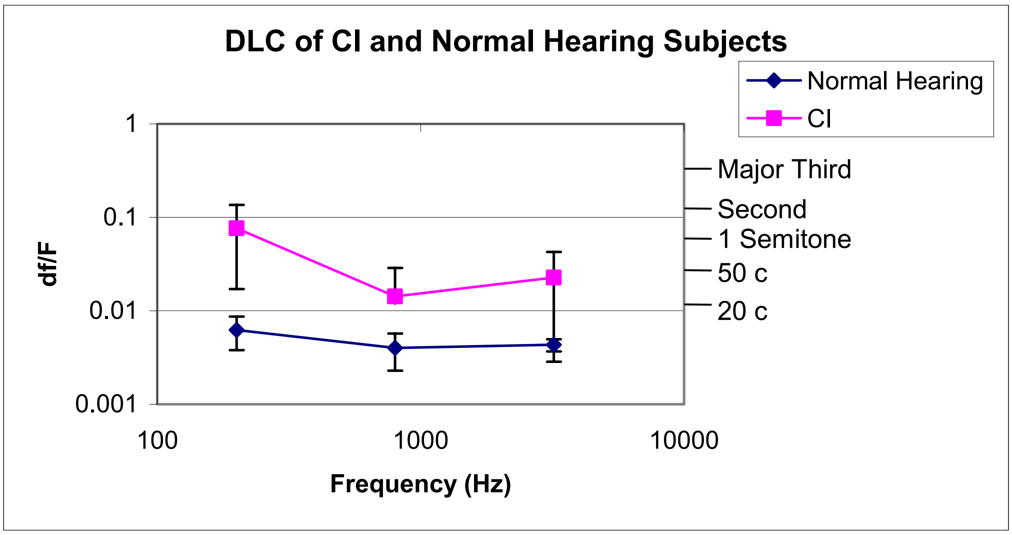


Figure 7-6 Frequency difference limens for CIR and NH participants

7.5 Discussion

The output of the piano consists of multiple frequencies and therefore its harmonics will stimulate more electrodes than a pure tone. This probably allows small changes in pitch to be determined from the harmonics that fall in regions of the cochlea where there are more electrodes/octave.

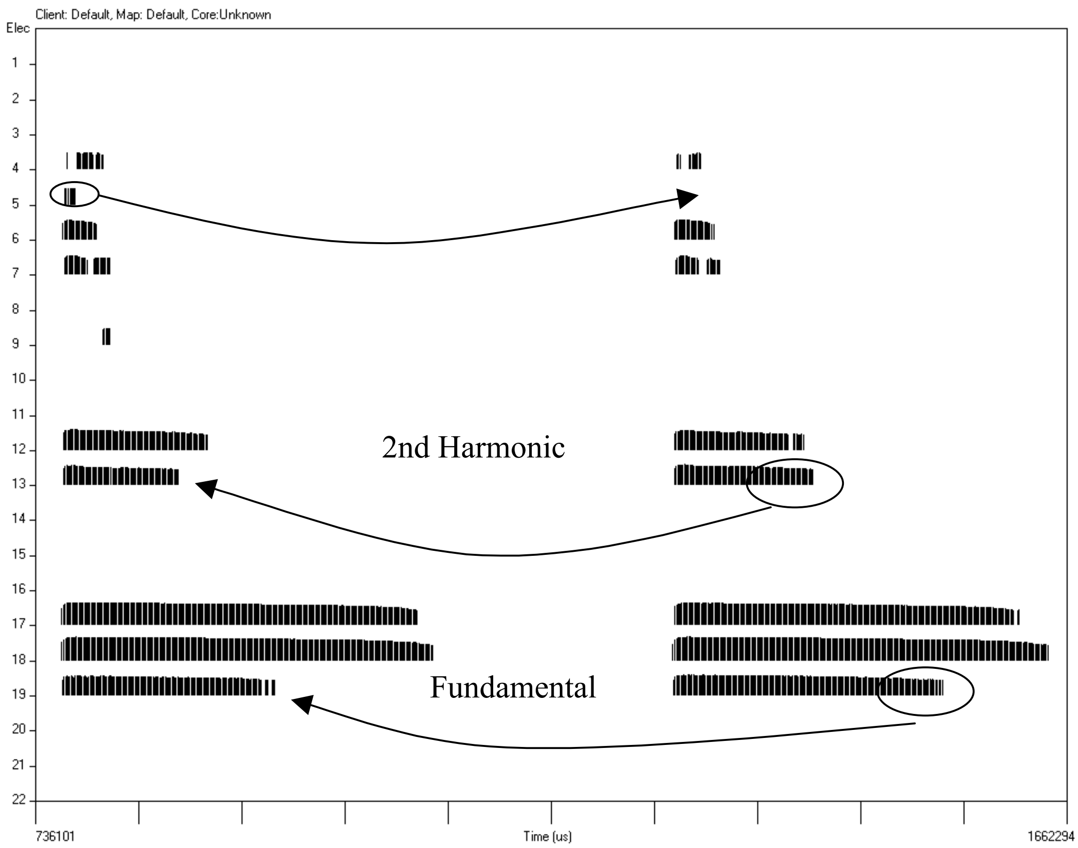


Figure 7-7 Electrodogram of two piano tones differing in pitch by the smallest JND at 800Hz using the ACE strategy (x tick is ~0.1s).

The electrodogram in Figure 7-7 is a capture of two piano tones that differ in frequency by the smallest JND at 800Hz when subjects were using the ACE strategy. Superimposed onto the electrodogram are circles indicating the differences between the representations of the two tones. The main differences are that different electrodes are stimulated as a result of the change in pitch used and the place pitch percept provides the difference.

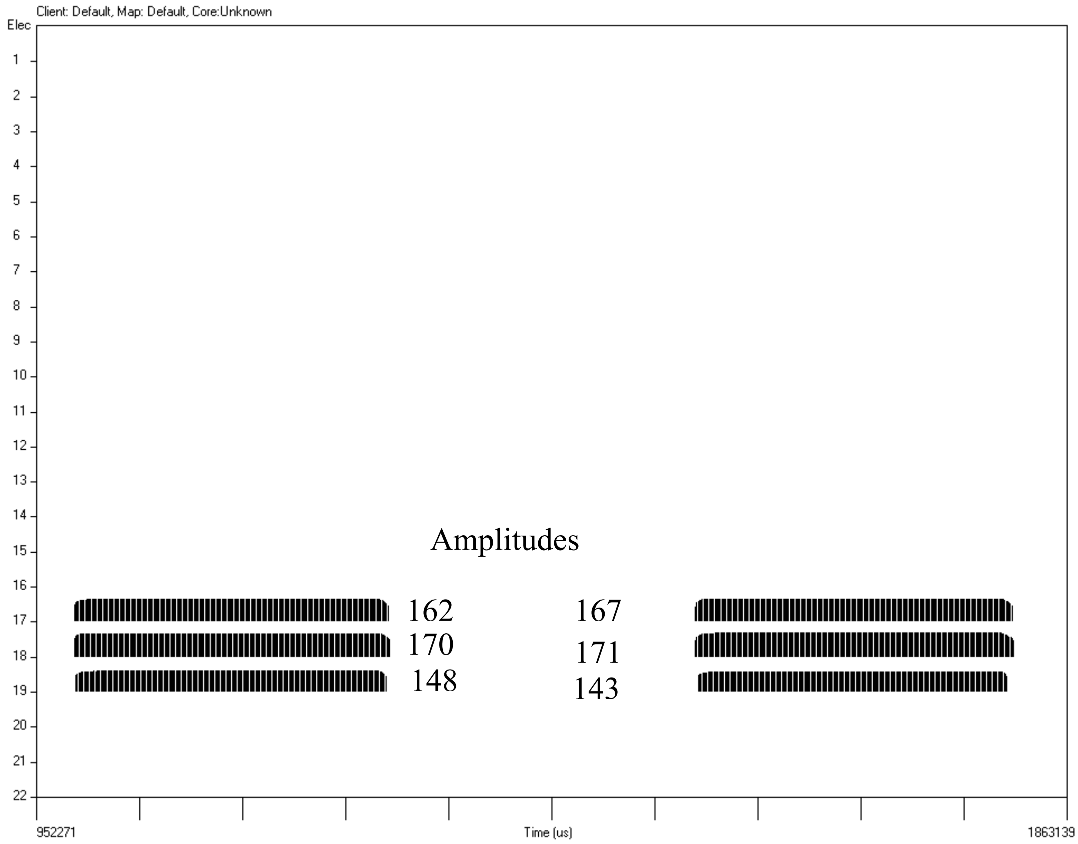


Figure 7-8 Electrodogram of the smallest JND at 800Hz for the pure tone stimulus using the ACE strategy (x tick is ~0.1s).

The electrodogram shown in Figure 7-8 is a captures of two pure tones that differ in frequency by the smallest JND at 800Hz when subjects were using the ACE strategy. The pure tone has no harmonics and thus only stimulates three electrodes (due to overlapping filters). In this case the same electrodes are stimulated and the difference that is detectable is the relative amplitudes as quoted the figure. This figure illustrates the capacity of CIRs to resolve subtle effects in the place domain.

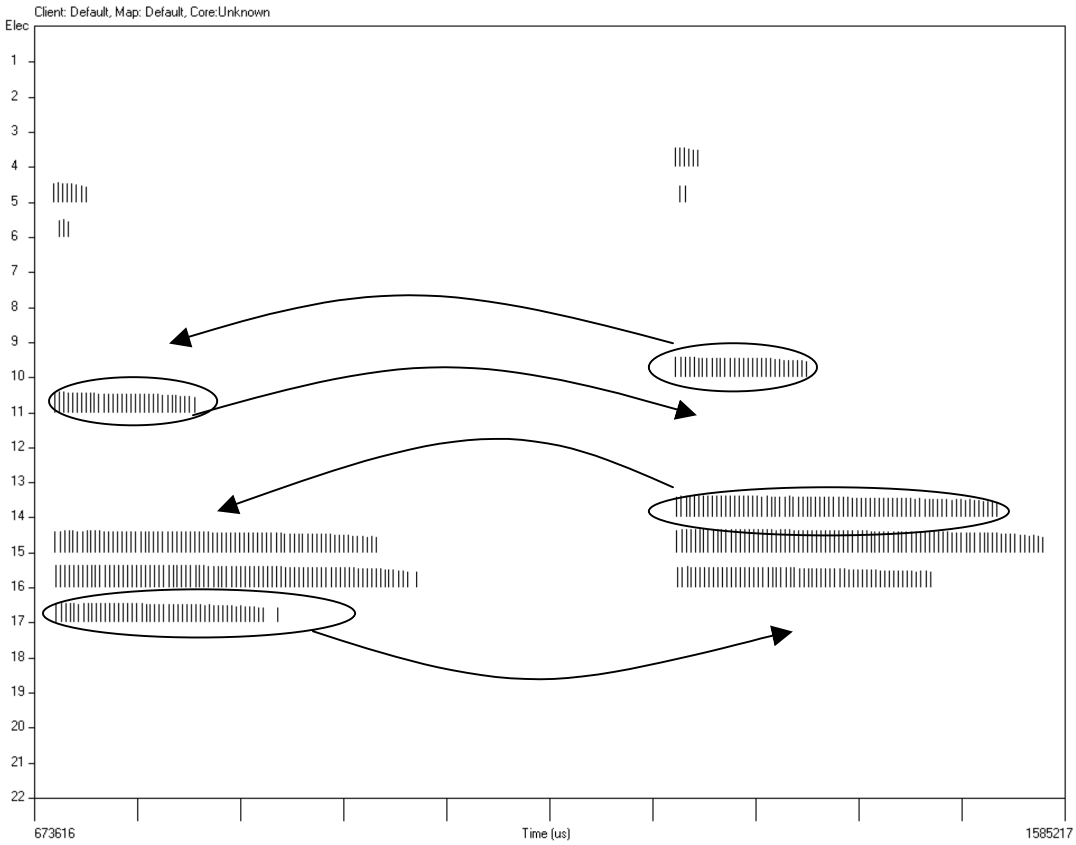


Figure 7-9 Electrodogram of the two piano tones that differ in frequency by the smallest JND at 800Hz using the SPEAK strategy (x tick is ~0.1s).

Using the SPEAK strategy, where the stimulation rate is much less, the recipient requires greater differences in frequency (above 800Hz) to identify the change than when using the ACE strategy. The differences (circled) that can be noted on the electrodogram provide the user with information to perceive the stimuli as different.

The shape of the filterbank used in current speech processor designs in which less resolution is available at lower frequencies appears to be responsible for CIRs' reduced discrimination ability of pure tones at lower frequencies. Above 800 Hz these differences diminish. The DLC and DLF for the pure tone appear to be of similar order and are most similar, albeit much worse than NH listeners. When comparing the DLC and DLF for the piano stimulus it was found that the ability to reliably detect the correct directional change in pitch, DLF, was much worse. The results found through these experiments hold promise for improving music perception by improving the perception of the direction of changes in pitch. It appears two courses of action can be taken to improve music perception in as much as improving correct directional changes in pitch.

Firstly, the production-perception feedback mechanism that exists for music is virtually non-existent for one who does not study music but merely listens to music. Similar feedback mechanisms that exist for speech helps CIRs quickly adapt to the new sounds they are hearing and correct for any mismatch of sounds they have. This for example could be hearing the consonants ‘p’ and ‘t’ as reversed: an error that would be quickly corrected through context. When the implant recipient listens to music there is no formal mechanism to correct for any mismatch of pitch or to help relearn pitches and instruments, as responses to music are more often emotive than a grammatical analysis. The only feedback that does exist is the memory of the CIR. However, most CIRs have gradually lost hearing so this memory of music is often reduced.

It is proposed that formal feedback be introduced into the rehabilitation of CIR to help relearn the sensations and stimulations that now represent pitch, harmony, and melody. This formal feedback will help match the new information with their memory for components like direction of pitch, and perhaps where pitch reversals occur can be re-learnt. This may seem ambitious but is similar to that which happens informally with speech.

The second proposal to improve music perception is to change the signal processing schemes to produce stimulation patterns that match what would have been heard with NH. This may remove the ambiguity and/or pitch reversals if the problem lies with obvious error in stimulation patterns. Pitch reversals that are due to neural pathways may be more difficult to overcome, but with careful measurement might also be accounted for.

7.6 Conclusions

The major result of this chapter is the result that the difference limen for change is better using the piano, but the difference limen for frequency is much worse. This result helps explain why CIRs report music sounds so poor in real life situations. They can perceive a change reasonably accurately but either the perceived pitch direction is in the opposite direction or the pitch direction is ambiguous.

There does appear to be an advantage of using the ACE over the SPEAK strategy especially above 200Hz in reducing the difference limen for change. When comparing the difference limens, both DLC and DLF, between the NH and CIR groups, the disparity was dramatic as expected. The NH group achieving up to nearly an order of magnitude better than the CIR group.

Chapter 8
Measuring Pitch Discrimination Using Alternative Techniques

> Yet it was impossible for me to say to people, "Speak louder, shout, for I am deaf." Ah, how could I possibly admit an infirmity in the one sense which ought to be more perfect in me than others, a sense which I once possessed in the highest perfection, a perfection such as few in my profession enjoy or ever have enjoyed.
>
> Ludwig van Beethoven
> Heiglnstadt
> October 6th, 1802

8.1 Introduction

The tests of chapter 7, while being thorough, are time consuming. They can take 3-4 sessions of up to two hours each. The researcher is confronted with the problem of wanting a short test that may provide much of the information provided in Chapter 7 but in a more compact form. In the speech-testing field, sentence and word lists are used to readily obtain speech perception levels in a single test session. The subject or researcher often does not have the time for multi-week evaluations to measure a level of pitch discrimination. This chapter reports measurements of the ability of CIRs, to discriminate pitch in tests that required less time to perform. For a more detailed analysis, the tests of chapter 7 can always be used, but if faster estimates are required, tests described in this chapter can be used. To provide a comparison the results of NH subjects under analogous conditions are also reported.

The most important features of any test is its ability to measure and differentiate. The absolute result may not be as important as the ability to distinguish either between the ability of two participants or between the same participant under two different conditions. Practically, the ability to differentiate between subjects is not as important as the ability to measure change in the same participant. This can be either through improvement with the same strategy, perhaps through training, or improvement when using a different strategy. Tests that can be administered quickly and can measure these changes are very useful.

The experiments reported here were designed to measure pitch discrimination using two different techniques:

i. Pitch Discrimination Test, ('PDT') - three octave version.
ii. Adapted Primary Measures of Music Audiation, ('PMMA') - Tonal subset.

Both use psychophysics software developed by the author for this study.

8.2 Pitch Discrimination Test

8.2.1 Introduction

The Pitch Discrimination Test (PDT) is used to gather a 'general score' for pitch discrimination ability. This test was devised to measure pitch discrimination ability over three octaves using a real instrument sound and/or a pure tone. The author wrote the software for this test using Microsoft's Visual Basic and a screen shot can be seen in the Figure 8-1.

The test runs on a personal computer (PC) and is controlled by the test administrator or is self-administrated if the participant is able to control a computer mouse necessary to select the response buttons on the computer screen. The stimulus is presented via a sound field at 70-85dB SPL.

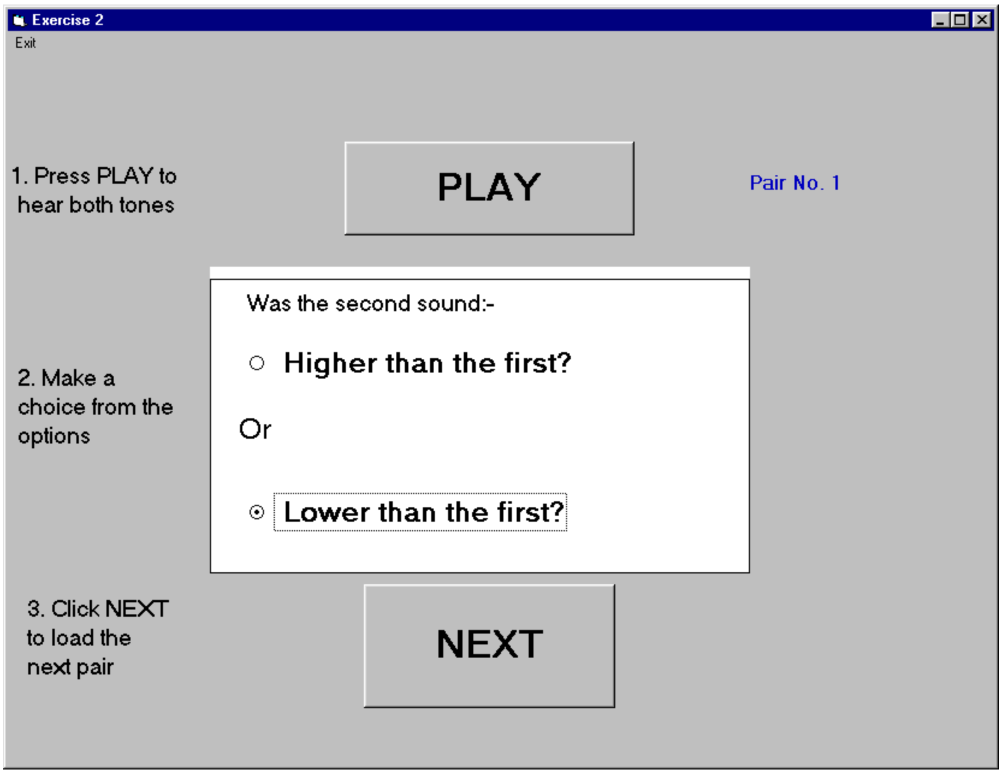


Figure 8-1 Screen shot of the PDT test

8.2.2 Aims

Is pitch discrimination level dependent on the complexity of stimulus?

This experiment uses two types of stimuli. The first is the pure tone (or sine wave) and the second is the piano. The reasons for this choice have been discussed in Chapter 7. When a complex stimulus is used, the peak-picking algorithm selects the largest harmonics of the signal and stimulates the corresponding sites on the electrode array.

Is pitch discrimination level dependent on speech processing strategy used by the CIR?

The software strategy that operates on the external speech processor is responsible for deciding which electrodes to stimulate, based on the incoming sound signal. There are several different categories of strategies, which include ACE, SPEAK, CIS and Analogue, which have been discussed in Chapter 2. These strategies have been developed primarily for the extraction of speech features rather than for music perception. This study compares pitch discrimination level for different categories of speech processing strategy.

Can the pitch discrimination level be improved with training?

When a CIR is first reintroduced to the hearing world, the feedback mechanism for understanding speech is immediate. This feedback mechanism is one of replaying the stimulus until comprehension is gained. For example one of the first sentences spoken to a CIR at 'switch on' is "Can you understand me?" The recipient will respond if they understand, otherwise the question is repeated, said more slowly etc until the meaning is conveyed effectively and the recipient can 'hear' the speaker. This learning process is continuous and an automatic part of retraining to understand speech with the new stimulation. Indeed ordinary conversation, among NH listeners includes many feedback mechanisms that regulate speech, to ensure that it is comprehended and which function as training in comprehending speech.

The understanding and comprehension of pitch and music however (both for CIRs and in general society) has a very weak feedback mechanism. When presented with a piece of music or a sequence of tones, the response is usually an emotional evaluation as opposed to an explicit understanding and remembering of the sequence of notes. There is little feedback to the CIR for relearning the direction of pitch that is heard by a NH person. A hypothesis was proposed that suggested that the CIR was not achieving their full potential when it came to music and pitch discrimination because of the lack of a feedback network. The effect on pitch discrimination of learning via a feedback network was examined.

8.2.3 Method

The test items are two one-second tones separated by one second of silence. The two types of stimulus to be tested are the sampled piano and pure tone (sine wave); both stimuli are controlled using the Musical Instrument Digital Interface (MIDI) protocol. The harmonics present in the stimulus represent relative strengths according to the acoustic piano timbre. A range of thirty-five semitones (almost three octaves) was used which has fundamental frequencies ranging from D3 (146 Hz) up to C7 (1046 Hz). A sine wave, with a 10% attack and decay was also used with some participants for comparison.

The task is a two alternate forced choice (2AFC) interval test. The tones separated by silence are presented to the participant who must decide whether the second tone is higher or lower in pitch than the first. CIR participants are instructed to adjust their processor to a comfortable loudness level. The initial interval is set to twelve semitones. The initial tone of the pair presented and direction of the interval is chosen randomly. An adaptive algorithm is used rather than exhaustive techniques such as methods of constant stimuli as the former reduces the time taken to find the participants minimum pitch resolution (Jesteadt, 1980). The algorithm is adaptive to the participant's responses and is a variation of the staircase method (Cornsweet, 1962; Truetewin, 1995). To reduce the interval size presented, the participant must achieve at least nine correct responses out of eleven presentations. This achieves statistical significance at a minimum of $\underline{p}<0.05$ level ($p=0.0327$). If there are three incorrect responses for a certain interval level, the interval size is increased. The length of the tests varies according to each participant's response. The minimum number of possible presentations is fifteen. The maximum number of presentations is fifty-five. The average number of presentations, calculated empirically, is close to forty, which requires approximately ten minutes to complete.

The algorithm has a memory of the last three interval levels the participant has achieved and adapts to the participant's response. Generally, if a participant responds correctly to an item, the interval size is reduced, based upon previous interval levels tested. Likewise, when the participant responds incorrectly, the interval size is increased. The algorithm continues until it finds an interval size judged significantly correct and an interval size one semitone smaller than this judged incorrect. The pitch discrimination score for that participant would be then the smallest interval judged correct. This test was specifically designed for CIRs and a minimum limit of the test is set to one semitone.

The author wrote a second version of the test software that provided the following three features:

i. The software provides a piano keyboard on the screen that allows the participant to press the piano keys by selecting them with the computer mouse. This enables the participant to try to find the notes after they are presented to reinforce the relationship of higher/lower with the graphical display of the keyboard. This can be seen in Figure 8-2.

ii. The software gives a 'correct/incorrect' response after the participant has selected the 'higher/lower' choice. This enables the participant to associate what they thought was a pitch direction with the real pitch direction.

iii. Once the 'higher/lower' choice is made, the software allows the participant to play the tones again and see them on the graphical piano keyboard. As each tone plays the corresponding key changes colour as can be see in Figure 8-3. This visual feedback reinforces the sense of high and low to relearn the perceived stimulus with an accepted version of pitch direction.

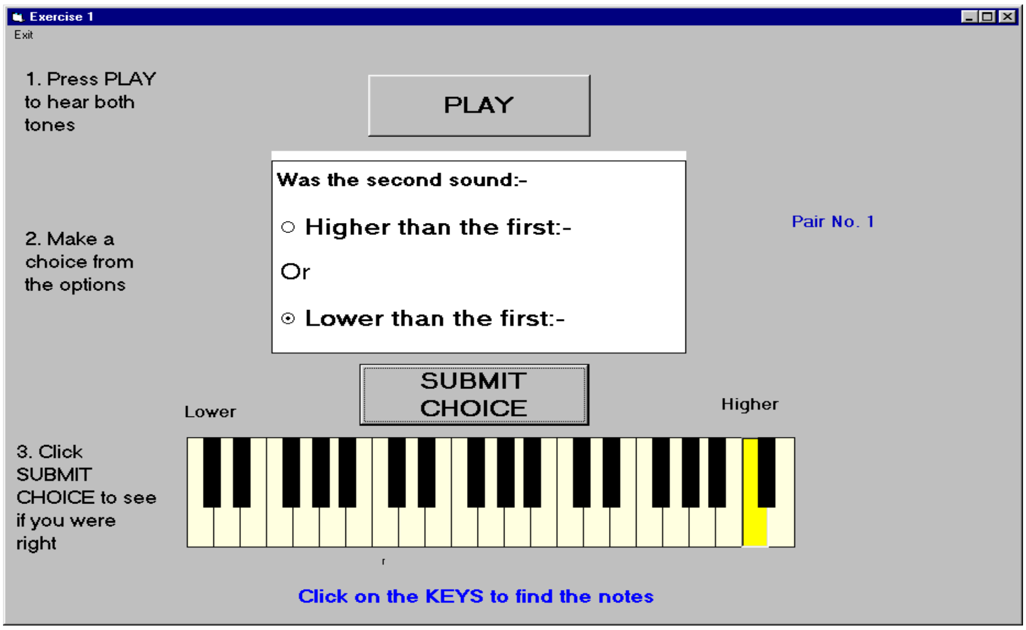


Figure 8-2 Screen shot of a participant using the keyboard to find the presented note.

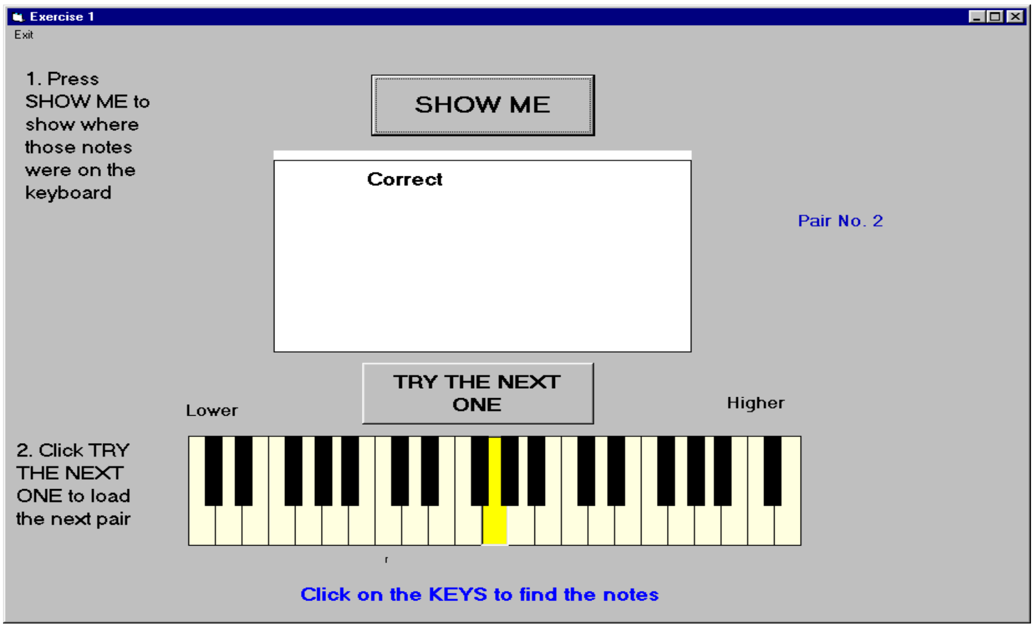


Figure 8-3 Screen shot of a participant being graphically shown where the presented notes were on the keyboard.

Twenty-four CIRs took part in thus study. Participant details are shown in Table 8.1. The average age of the recipients was 60 years and 58% were female. All participants were postlingually deafened and had greater than six months (average 91 months) experience with their present device, except for participant CIR8 who has only one month experience. Twenty NH adults also took part in this experiment for comparison. The NH population who participated in the study ranged in age from 13 to 55. 16 of the participants were from the University of Iowa's Department of Music and Department of Otolaryngology and most had musical training. 80% of the participants were female.
To increase the number of recipients taking the test, seventeen of the twenty-four subjects (CI5-CI22) were tested by Kate Gfeller at the University of Iowa in the USA in collaborative research.

Participant	Sex	Age	Device	Strategy	Experience (mo)
CIR6	F	41	CI24M	ACE	6
CIR7	M	63	CI24M	ACE SPEAK	36
CIR8	M	47	CI24M	ACE	1
CIR9	F	59	CI22M	SPEAK	48
CIR10	M	74	CI22M	SPEAK	60
CIR11	F	40	CI22M	SPEAK	50
CI5	F	59	Ineraid	Analogue	108
CI6	M	75	Ineraid	Analogue	126
CI7	F	61	Med-El	CIS	126
CI8	M	48	CI22M	SPEAK	112
CI9	F	69	Med-El	CIS	126
CI10	F	72	CI22M	SPEAK	106
CI11	M	44	Med-El	Analogue	102
CI12	F	73	CI22M	SPEAK	144
CI13	M	38	CI22M	SPEAK	124
CI14	F	52	CI22M	SPEAK	105
CI15	F	41	Clarion	CIS	60
CI16	F	78	Med-El	CIS	138
CI17	F	69	CI22M	SPEAK	92
CI18	M	63	CI22M	SPEAK	121
CI19	F	74	CI22M	SPEAK	96
CI20	M	80	Med-El	CIS	144
CI21	M	78	CI22M	SPEAK	104
CI22	F	51	Clarion	CIS	51

Table 8.1 Participant details

8.2.4 Cochlear Implant Recipient Results

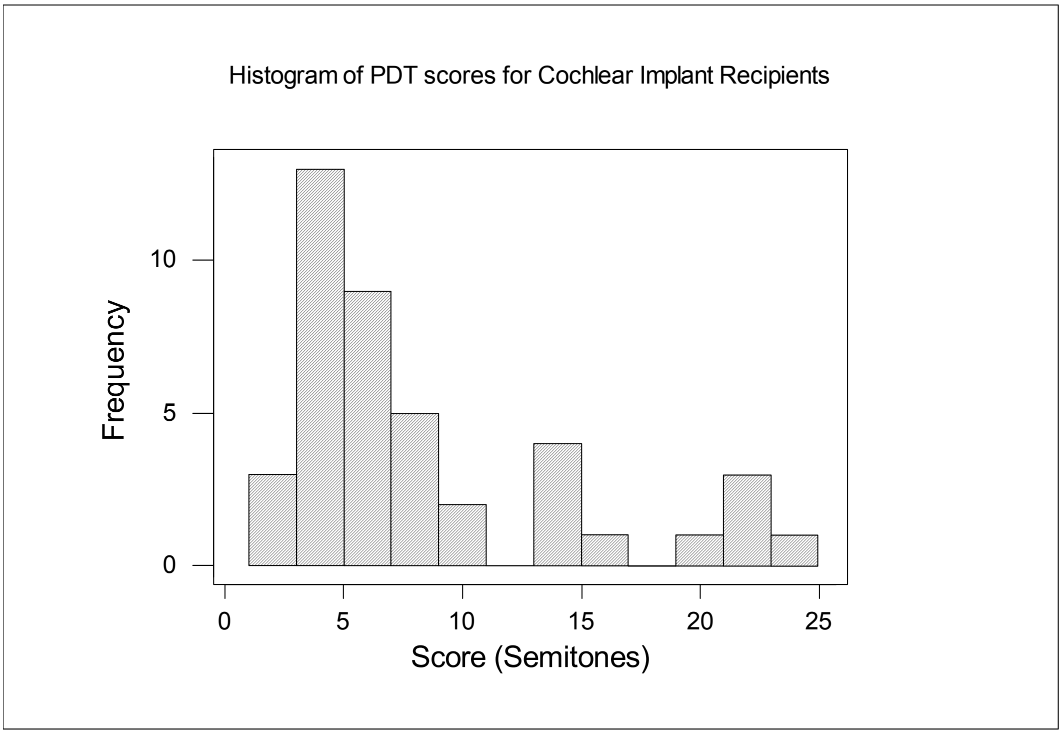


Figure 8-4 A histogram of PDT scores (in semitones).

The results are shown in Figure 8-4 in the form of a histogram. The mean score of the CIR was 7.905 semitones with a SD of 6.172. It can be seen that the majority of the CIRs score below 10 semitones while a few subjects score around 15 and another small group around the 2-octave region (24 semitones).

Is pitch discrimination level dependant on the complexity of stimulus? (25 participants)

An ANOVA was performed on the results and although the mean of the PDT score when participants are presented with a pure tone (mean 5.385, SD 3.203) is less than when presented with the stimulus of a piano (mean 9.25, SD 6.80), the result was not significant (p= 0.126).

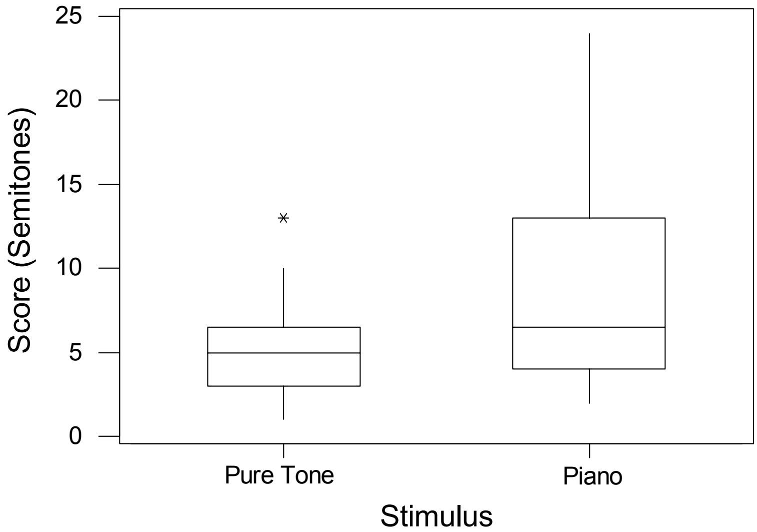


Figure 8-5 PDT score in semitones grouped according to stimulus type.

The situation does arise with a CIR that a peak-picking algorithm can actually present information that may mislead the CIR in the nature of the pitch. The electrodogram, Figure 8-6, is a capture taken in response to a two octave piano scale. The regions that are circled should represent an increase in pitch as time progresses from left to right on the graph which can be observed by the shift of lower frequency stimulated electrodes (higher number) to higher frequency electrodes (lower numbered). In the left-most circle, the stimulations on electrode 13 is replaced by a stimulation that will portray a lower frequency on electrodes 14 and 15. In the right-most circle, even though the new stimulus on electrode 17 would be perceived as in increase in pitch, the new stimulation on electrodes 12 and 13 and lack of stimulation on electrodes 8-11 could represent a decrease in pitch, resulting in ambiguous change in direction of pitch and would be mostly likely perceived as not a change in pitch but a change in timbre or type of instrument, as is often reported. Some subjects actually report that the signals do not just differ in pitch but report that “part of it goes up and part of it goes down” (personal communications). This ambiguity of direction of pitch change due to the peak-picking algorithm most likely results in the inability to correctly track changes in direction of pitch.

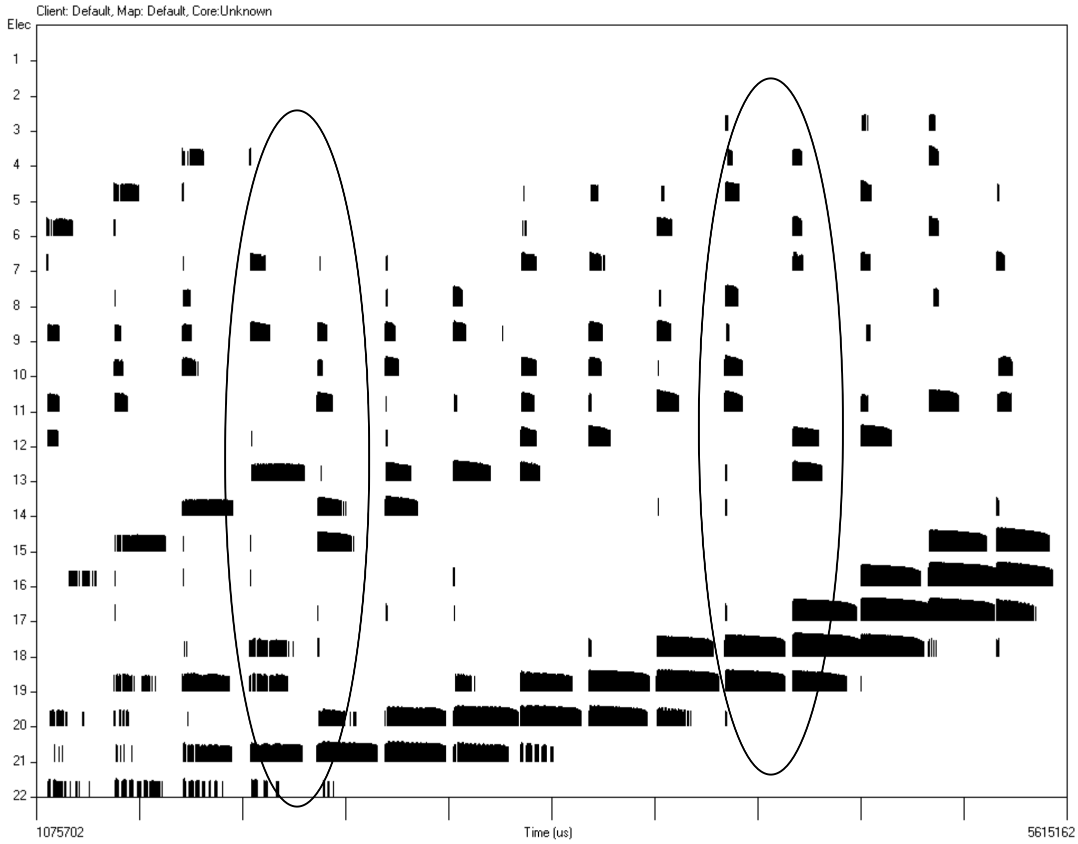


Figure 8-6 Electrodogram showing how a peak picking algorithm can present ambiguous information about the pitch (x tick is ~0.45s).

Is pitch discrimination level dependant on speech processing strategy used by the CIR?

An ANOVA was performed on the data and the effect due to the speech processing strategy type used was examined. The difference in the PDT score due to strategy type was significant at the $p<0.05$ level ($p=0.015$). Upon further investigation, using Tukey paired comparisons, no individual effects were found (Informally, a trend can be observed in that the mean score of the ACE strategy users was better compared to the mean score of the Analogue strategy users but was close to, but not significant $p=0.0515$).

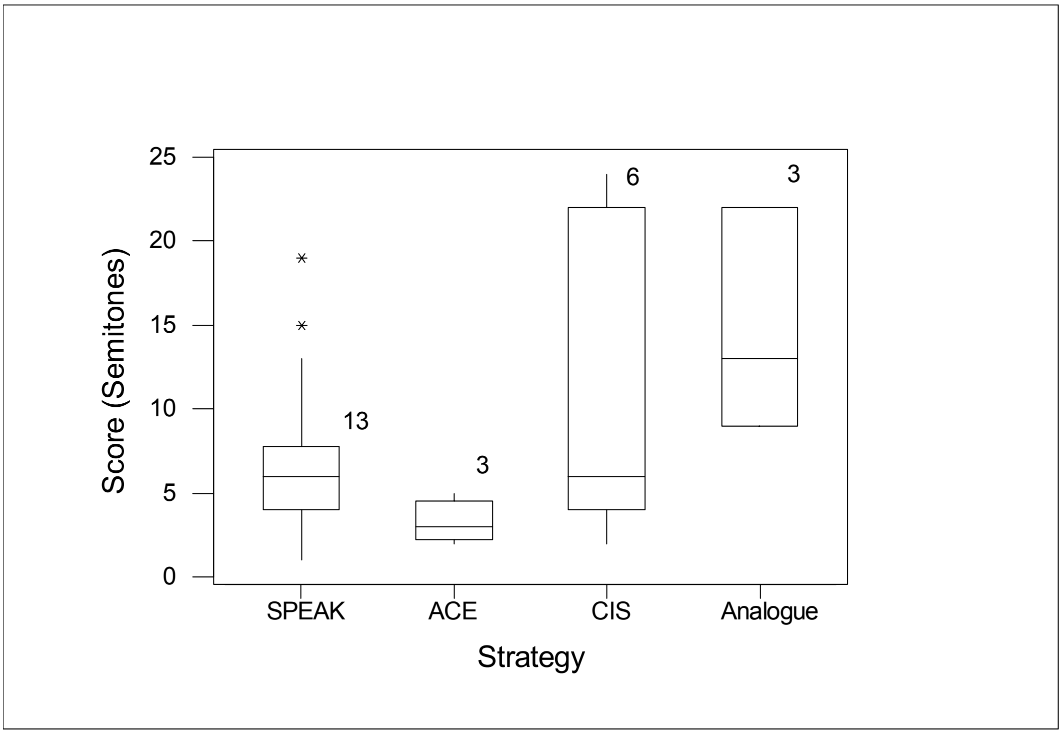


Figure 8-7 PDT scores grouped by speech processing strategy. The number of participants in each group is shown next to each box. Subject details can be found in Table 8.1

Can the pitch discrimination level be improved with training?

An ANOVA was performed on the data and the author examined the effect due to test type. Although the mean scores for the PDT with training/feedback were lower (mean 4.80, SD=4.60) compared to the PDT without training/feedback (mean=8.81, SD=6.53) the difference in the means was not significant (p=0.197).

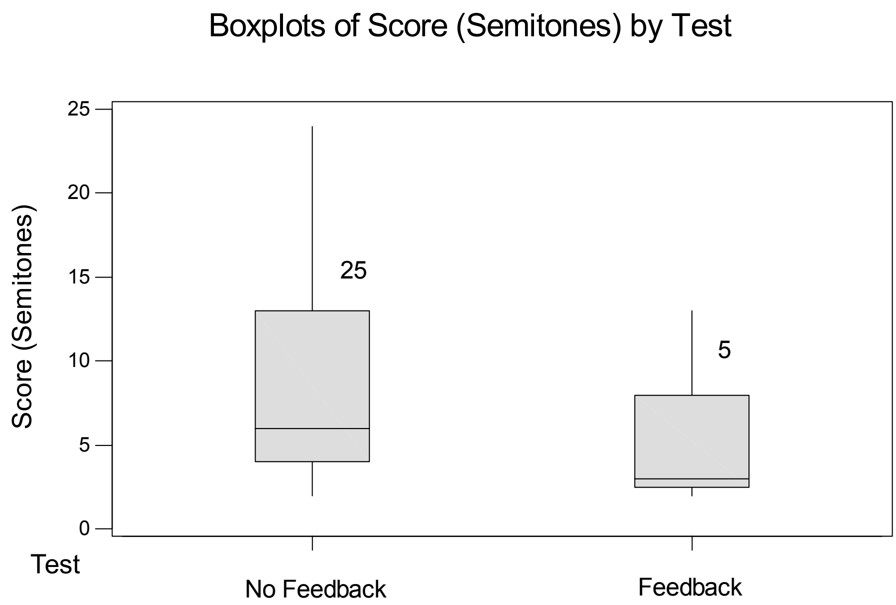


Figure 8-8 The PDT score when grouped by test type. The subject number that participated in each test is shown next to each box.

Further Tests

There was a slight negative correlation between semitone score and age (Pearson correlation of Score (Semitones) and Age = -0.213) but the result was not significant.

There was no correlation between semitone score and months usage (Pearson correlation of Score (Semitones) and Usage = 0.049).

8.2.5 Normal hearing results

The results of the CIR participants were compared to NH participants by performing an ANOVA on the data. The difference in the means was statistically different at the $p<0.001$ level. The mean score of the NH participants was 1.15 semitones, SD=0.3663. This score is limited by the minimum resolution of the test, which is one semitone. This limit was found to be suitable for CIR participants but obviously not for a real test of NH participants but does serve as a useful comparison. The mean score of the CIRs was 7.905 semitones with a SD of 6.172. The NH population age ranged from 14 to 55 (mean 20.5), 75% were female and 90% were involved in some way with music. This last factor may also be responsible for the good scores produced by the NH group.

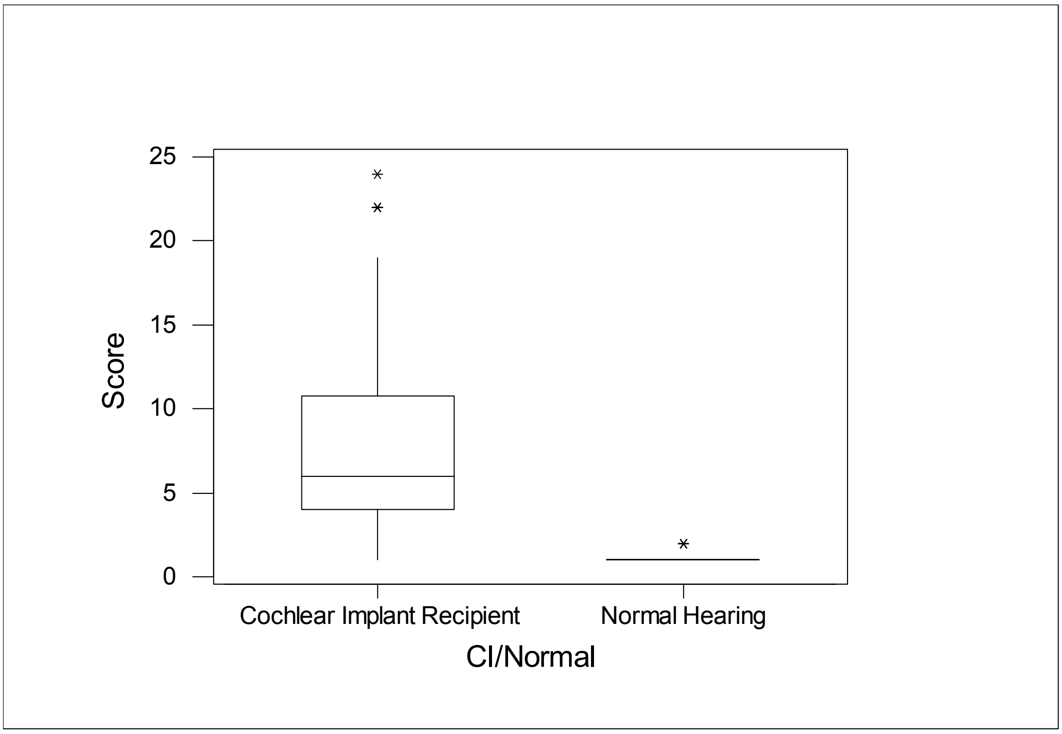


Figure 8-9 The PDT score when grouped by participant group.

8.3 Primary Measures of Music Audiation (PMMA)

8.3.1 Introduction

This study is designed to investigate perceptual accuracy of melodic patterns by CIRs. The test material used is the tonal subset of the Primary Measures of Music Audiation, PMMA (Gordon, 1979). This test is used in its normal auditory format. The author also transcribed the PMMA to MIDI format to enable changes to the stimulus timbre (instrument), and improved the presentation to be performed for comparative testing. In the normal auditory format verbal prompts are given to indicate the question number. The MIDI files used in this study have the advantage that they have effectively removed the verbal prompt, which was thought to cause distraction and complicate the task especially for CIRs (Gfeller and Lansing, 1992).

The PMMA is a widely used standard test of musical perception for use with children in kindergarten to third class. The PMMA consists of two components, tonal and rhythm, each of which is composed of 40 pairs of synthesised items. Each member of the pair is approximately 1.3 to 2 seconds long separated by 5- 6 seconds of silence. The participant is required to respond with ‘SAME’ or ‘DIFFERENT’ on a printed answer sheet to indicate their perception of differences in short tonal or rhythm patterns. The tonal test contains:

i. Thirteen two-tone pairs.
ii. Twenty-five three-tone pairs.
iii. One four-tone pair.
iv. One five-tone pair.

The pitch range used in this test is C4 (261Hz) to F5 (694Hz). The item pairs are intended to have the same rhythmic patterns although close examination reveals that some of the presented notes have a small gap between them, while others have no gap. There are no pairs that have a comparison between a set of notes having gaps, to a set that has no gaps so this should not affect the overall accuracy of the test. The rhythm test presents tones that are different in duration or intensity at the common frequency of C5 (520Hz).

The PMMA, although originally designed for kindergarten to third class children, has been widely used with a variety of participant groups including adults, elderly adults, disadvantaged and retarded children and also in one study using CIR participants. The study that tested CIRs with the tonal subset of the PMMA (Gfeller and Lansing, 1991) used two groups of participants using either the Nucleus™ F0F1F2 or the Ineraid 4 channel filter bank analogue speech processing strategies. Their results are included as a comparison in the results section 8.3.5. The F0F1F2 strategy presents formant amplitude F1 and F2 information at a stimulus rate determined by the F0, fundamental frequency. The analogue strategy divides the frequency in four filter bands and presents the amplitude on 4 electrodes using analogue stimulation. The study presented here is with subjects using more recent and advanced speech processing strategies like SPEAK and

ACE and thus it will be useful to observe whether, if advances in speech processing strategies have also resulted in advances in pitch discrimination skills as measured by the PMMA.

8.3.2 Aims

Is there a difference between the commercially supplied audio format of the PMMA test and the MIDI transcribed version of the test? Are there any differences in the scores achieved when the stimulus is a piano or sine wave?

The commercially supplied test by Gordon (1979) provides verbal prompts (e.g. 'first', 'second') and uses simple graphical symbols on the answer sheet and verbal prompts (e.g. 'apple', 'fork'). Gordon generated the stimulus using a Moog™ synthesiser. This type of presentation is suitable for NH children. For use with CIRs, it has been suggested that it would be less distracting for CIRs to have the verbal prompts removed. Another feature of the test that was thought to influence the results of the tests with CIRs was the nature of the stimuli presented. To test this theory a comparison was performed between the commercial test and a modified version that had no verbal prompts and used a piano or sine wave stimulus. This test was conducted in order to assess the validity of a modified test as well as measure the effect of the complexity of the stimulus on the PMMA score.

Is the PMMA a suitable test for measuring pitch discrimination ability with the current state of speech processing strategies and how do CIRs tested in this study compare to previous CIRs tested in a previous study who were using a different strategy?

A previous study that has used the PMMA with CIRs found the mean score to be around 77%. This suggests the scores in the PMMA might saturate at high levels if more recent speech processing strategies improve pitch discrimination ability. If that occurred, participants, strategies, implants and other important parameters could not be assessed with confidence. Considering the strategies such as F0F1F2 that were used by CIRs in the previous study were introduced around 1985 (Tye-Murray et al., 1990), it is useful to reassess the suitability of this test for use with current speech processing schemes such as Speak and ACE. If the results obtained are getting close to or surpassing 80% it would suggest a more difficult test might be needed. It remains to be seen whether the newer coding strategies that have improved speech perception have also improved pitch discrimination, as tested by the PMMA.

What were some of the common errors made by CIRs when performing a task like the PMMA? Do the adult CIRs make errors similar to those made by children? Are there trends in the types of errors made by CIR taking the PMMA test?

The PMMA test involves judging whether two sets of stimuli the same or different. In the example Figure 8-10, the two sets of three notes would be correctly judged as 'different'.

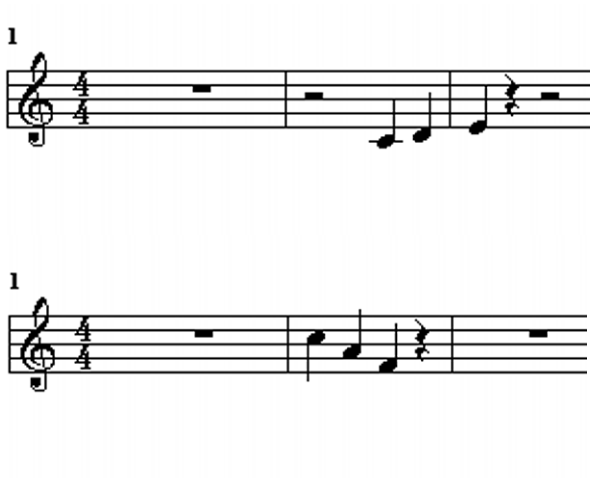



Figure 8-10 An example from the PMMA – tonal test

There are several types of presentations; in some, only one note in the two phrases is different; in others the phrase may be inverted. If there were trends in the types of mistakes made, it might lead to ways in which speech processing strategies could be improved or ways in which training could be given to CIR's.

How do adult CIRs compare to the NH children reported in the literature for whom the test was designed?

Gordon (1979) designed this test for use with kindergarten to third grade. Gordon conducted this test on approximately one thousand children. The results from the present study can thus be compared with the study conducted by Gordon in order to assess the CIR participant's abilities and limitations.

Do the different types of error give information about pitch resolution?

Several skills are tested in the PMMA. One is the ability to recognise different pitches, which would be tested by phrases that have one note difference. Another skill tested is that of remembering the time sequence in which they occurred. This skill would be measured by ascending and descending phrases or by an inversion of a phrase. If the only limit a CIR participant has, is the resolution of pitch due to processor/electrode/surviving nerves limitations, than this may become noticeable in the statistics of phrases that they incorrectly judge different or incorrectly judge the same.

Is it possible to predict the results obtained from examining the resolution of the filters as they vary in frequency as used in the speech processors of CIRs?

Using the CIR participant's filterbank, which is a major component of the speech processing strategy, it may be possible to predict from visually inspecting electrodograms or statistically analysing their output what limits the devices will have in conveying pitch which would limit the participant's minimum pitch discrimination threshold. When a new signal-processing scheme is introduced it could then be seen whether it delivers more pitch related information and increase the capabilities of the CIR participants by a reduction (an improvement) in their pitch discrimination scores.

8.3.3 Method

These experiments only used the tonal subset of the PMMA. This was because previous studies (Gfeller and Lansing, 1992) reported that while the tonal test differentiated CIR participants well, scoring them within the 50-80% difficulty range, the rhythm test, however, lacked the range of difficulty to differentiate participants, with the average score being 85%, outside of the preferred range. (Technical Bulletin cited in Gfeller and Lansing, 1992). The speech processing strategy that was tested in this previous study was the F0F1F2 strategy. The more recent speech processing strategies such as SPEAK and ACE have not been tested using this sort of test and thus will be tested here and compared against the results from the literature. The tonal PMMA test is useful, as it does not require musical training or nomenclature skills. It can be administered within 40 minutes (20 minutes for the tonal component) and the literature provides comparative data for normal children (Gordon 1979), hearing impaired adults (Darrow, 1987), mentally retarded adults (Braswell et al., 1988), the elderly (Gibbons, 1983a; Gibbons, 1983b) and CIR adults (Gfeller and Lansing, 1992; Gfeller and Lansing, 1991).

8.3.4 Participants

Two participants took part in the tests involving the PMMA. Participant details can be found in Table 8.2.

Participant	Age	Implant type	Processing Strategy	External processor
CIR6	41	N24	ACE 8/1800	Sprint
CIR7	63	N24	ACE 8/1800 SPEAK	Sprint Esprit

Table 8.2 Participant Details

8.3.5 Results

Participant 1. CIR6-ACE

Raw Audio	MIDI-Piano	MIDI Sine	Mean
75%	80%	80%	78%

Participant 2 CIR7- ACE/SPEAK

Raw Audio-Speak	MIDI Sine-ACE	Mean
80%	80%	80%

Table 8.3 PMMA results

There does not appear to be any difference between the commercial supplied audio format and the MIDI format of the tests, however with a small number of participants it is difficult to be confident of this conclusion. The participants commented that the MIDI version of the test was easier to listen to since they just had to listen to the notes and found the recorded voice distracting, and unnecessary. It was therefore concluded that the MIDI version was easier to use and also gave comparable results. It would however be harder to test with children as it is possible children, with lower concentration ability than adults, would need the verbal prompts, such as in the original version of the test, to help them keep place during the test. There appears to be no difference in mean scores when the stimulus is a pure tone or piano. This supports the results of Chapter 7 of difference limens for change with limens for piano and pure tone being similar.

The technical bulletin cited in (Gfeller and Lansing, 1992) suggested tests should have participant scores in the range of 50-80% for moderate test difficulty. The average score of the subjects used in this study was 77.5%, which suggests that this test is just within the range required to measure the pitch discrimination ability and may be approaching a ceiling. Further research is required into the suitability of a similar but more advanced test such as those Intermediate and Advance Measures of Music Audiation, the IMMA and the AMMA both by Gordon. Table 8.4 contains the results of this experiment as well as the results of Gfeller and Lansing (1992). The results suggest that any improvements in speech perception as made available by more recent speech processing strategies such as ACE and SPEAK are not reflected in improvements in scores from the PMMA test. This could mean one of two things. Firstly, that the improvements in speech perception found with newer coding strategies do not result in improvements of pitch discrimination scores or secondly that the improvements are not found with this test due to possible ceiling effects.

	This Study	**(Gfeller & Lansing, 1992) –Average**	**Nucleus**	**Ineraid**
Mean	77.5%	77%	80%	76%
SD	9%	10%	8%.	12%

Table 8.4 Results from this experiment and (Gfeller & Lansing, 1992)

The most common mistake appears to be when the first notes of three note phrases are different by a tone. There are two items presented of this type, item 26 and 32. Another common mistake is when there is an inversion in the three-note phrase in which the notes again differed by one tone. This suggests that the limiting factor for the CIR is their ability to detect changes in pitch especially when they are one whole tone or less.

When comparing the results of grade three, NH children (Gordon, 1979) with those of this study, it was found the seven most difficult items for the NH children coincided with the six most difficult items for these CIRs, listed in Table 8.5. Although the nomenclature is confusing Gordon has called the percentage of participants who scored the item correct, ‘difficulty’ and thus the higher the difficulty value, the easier the item.

Item No.	Difficulty	CIR most difficult items rank	Type of Presentation
19	34	3	First note different
29	52	5	First note different
21	59	4	Inversion of 3 tones
26	64	1	First note different
24	65	8	Second note different
32	73	2	First note different
28	78	6	Same: Ascending 3 note scale

Table 8.5 Common difficult items rated by NH children(Gordon 1979) and CIRs (this study)

These results suggest that NH children and CIRs are making similar errors though perhaps for quite different reasons. The NH child has most likely not yet learnt the skill to differentiate small changes in pitch when part of a melodic presentation is different or when presented with an inversion of phrase. A child may not yet have the skill to remember the sequence of notes in the correct order but may be able to recognise that the same notes were presented. For the CIR it is quite likely that the resolution required to discriminate the small changes in pitch may be limiting the ability to differentiate between pitches and appears to be the major cause of error. If this is the case, one might expect more different items judged the same than similar items could be judged different.

Results from the literature for NH children (Gordon, 1979) when compared with the mean CIR score from this study of 78% would be aligned with the pitch discrimination ability of a child somewhere between a grade one (mean of 75%) and grade two (mean of 81%). The average score of a CIR is however, far greater than hearing impaired grade three children (mean of 58%) as reported in (Darrow, 1987). The children investigated by Darrow had moderate to severe hearing losses, used conventional personal hearing aids and had stimuli presented over headphones at sound levels above their threshold. These results found in this study are comparable to elderly “care home residents” reported in (Gibbons, 1983b) where the researchers found the mean score in the tonal subset to be

80% with a SD of 5.81%. The errors were thought to be due to participant's inabilities to discriminate small interval changes in melodies.

There are an equal number of same and different presentations in the PMMA test. The response form the CIRs in this study is skewed strongly to judging a pair the same when if fact they were different. 74% (32/43) of incorrect judgments were due to this error. This strongly suggests that the resolution required to differentiate small changes in pitch is not available to these CIRs.

8.4 Discussion

The results of the three octave PDT suggest that even though the pitch discrimination level found by this test was dependent on speech-processing strategy for the CIR group, the scores are still much poorer (larger) and have more variance than the NH group. The mean score of CIRs was close to 8 semitones, whereas the NH participant was around 1.15 semitones and this was limited by minimum resolution of the test being one semitone.

This PDT was designed to give an overall score of pitch discrimination measured over three octaves. It can be used as a single measure to compare strategies, implants and individual participants and is relatively fast to administer (~10 minutes for most participants), and is automated so it may be administered with little operator training. A criticism of this test is that it takes an average pitch discrimination value over a range of three octaves, even though it has been shown in Chapter 7, that the pitch discrimination is a strong function of frequency. The explanation for this is that this test provides a simple guide to a measure of the recipient's overall pitch discrimination level even though the underlying dependence on frequency may vary. If detailed information is sought, more rigorous but time consuming tests, such as the JND and PDT (isolated frequency) used in chapter 7, can be performed. One can compare this with speech testing where often an overall speech score is given even though there may be a greater variation among vowels or consonant, of different types, but where it is still useful to give an overall speech comprehension score.

Modification of the PMMA test material by converting it to a MIDI format has the benefit to CIRs of removing the verbal prompts and also allowing stimuli of different timbres to be used. The results suggest that the tonal subset of the PMMA is a approaching a ceiling limit with the test not being ideal for differentiating subject's abilities.

8.5 Conclusion

The PDT test is a fast method of obtaining a score that gives an indication of pitch discrimination ability.

The PMMA test showed no improvements were found using the ACE strategy tested in this study with the F0F1F2 strategy tested previously. This may suggest that, while speech perception has been improved with advances to speech coding strategies, pitch discrimination levels have remained unchanged.

Chapter 9
Music-L: A Novel Signal Processing Scheme For Improving Music Perception

9.1 Introduction

The research reported in this thesis is inspired by a clear practical need. The results reported in previous chapters are combined to determine important parameters in a new processing strategy that aims to improve music perception.

This chapter will discuss some of the limitations of current signal processing schemes for processing music. A new strategy, for which the concept was filed as an international PCT patent application entitled "Multi-Rate Speech Processor" on 13 July 2000 (Wolfe, Carter, Frampton, Parker & Fearn.) This new strategy Music-L will be discussed and its implementation described.

9.2 Current signal processing schemes

Current signal processing schemes include SPEAK, ACE and CIS, described in Chapter 2. A limitation for conveying detailed pitch information using these processing schemes is that, after the amplitude is extracted from each filter, the stimulation on each electrode is at a constant rate. This stimulation will elicit a place pitch percept but does not provide enough information to allow discrimination of small pitch differences (~ 8 semitones in pitch direction task) to allow the differentiation of simple melodies, let alone complex music or harmony. An electrodogram of a typical constant rate strategy is shown in Figure 9-1. Figure 9-2 shows how pitch ambiguity may occur using a typical peak picking constant rate strategy.

Figure 9-1 Electrodogram of a constant rate strategy: Pitch is elicited by place information (Electrode) and temporal amplitude modulations below 187Hz (x tick is ~0.015s)

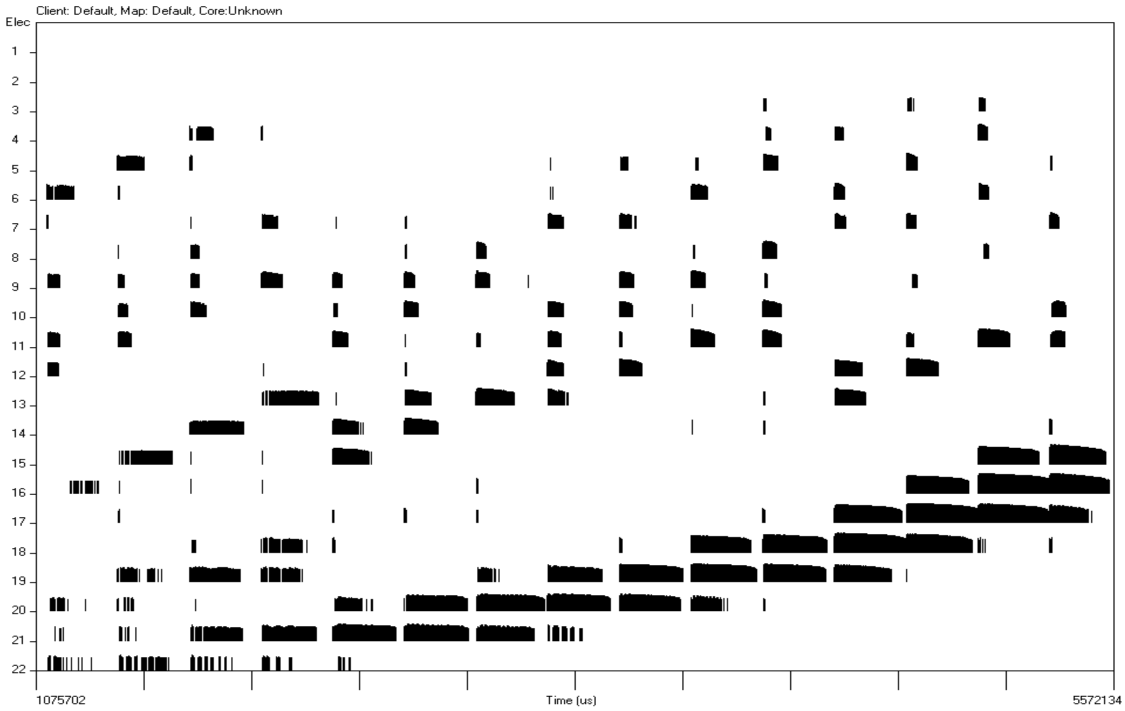


Figure 9-2 Electrodogram of possible pitch ambiguity: The tones shown are a rising scale but as shown may provide information that may sometimes mislead the CIR and be perceived as a decrease in pitch (x tick is ~0.45s).

9.3 Description of coding strategy

The Music-L processing scheme was implemented on the SPEAR (Speech Processor for Electrical and Acoustical Research). This device is a research processor that uses a Motorola 56309 Digital Signal Processor (DSP) developed by the Co-operative Research Center (CRC) for Cochlear Implant and Hearing Aid Innovation. The DSP is a 24 bit fixed-point device able to operate up to 80 Million Instructions Per Second (MIPS). Some of the code that performs the basic input, filtering and output are provided as core routines. All other code was written by the author to implement the new Music-L strategy. The Music-L scheme has two current versions, called Voc-L and Instrument-L, which are specified by:

i. Voc-L: Provides five channels below 1000Hz stimulating at the extracted frequency while all other channels stimulate at a constant rate of 1000pps. This strategy is intended to improve music perception while not degrading speech perception.

ii. Instrument-L: Provides ten channels below 1000Hz stimulating at the extracted frequency while all other channels stimulate at a constant rate of 1000pps. This strategy is intended to improve music perception but may degrade speech perception due to the different filterbank used.

A comparison of important parameters for the two new strategies and ACE are shown in Table 9.1.

	Voc-L	Instrument-L	ACE
Sampling Rate	16KHz	12KHz	16KHz
FFT bin width	125Hz	93.75Hz	125Hz
Number of Rate Channels	5	10	0
Number of Channels below 1000Hz	5	10	5

Table 9.1 Comparison of the Voc-L, Instrument-L and ACE strategies

A System block diagram of the Music-L signal-processing scheme is shown in Figure 9.3.

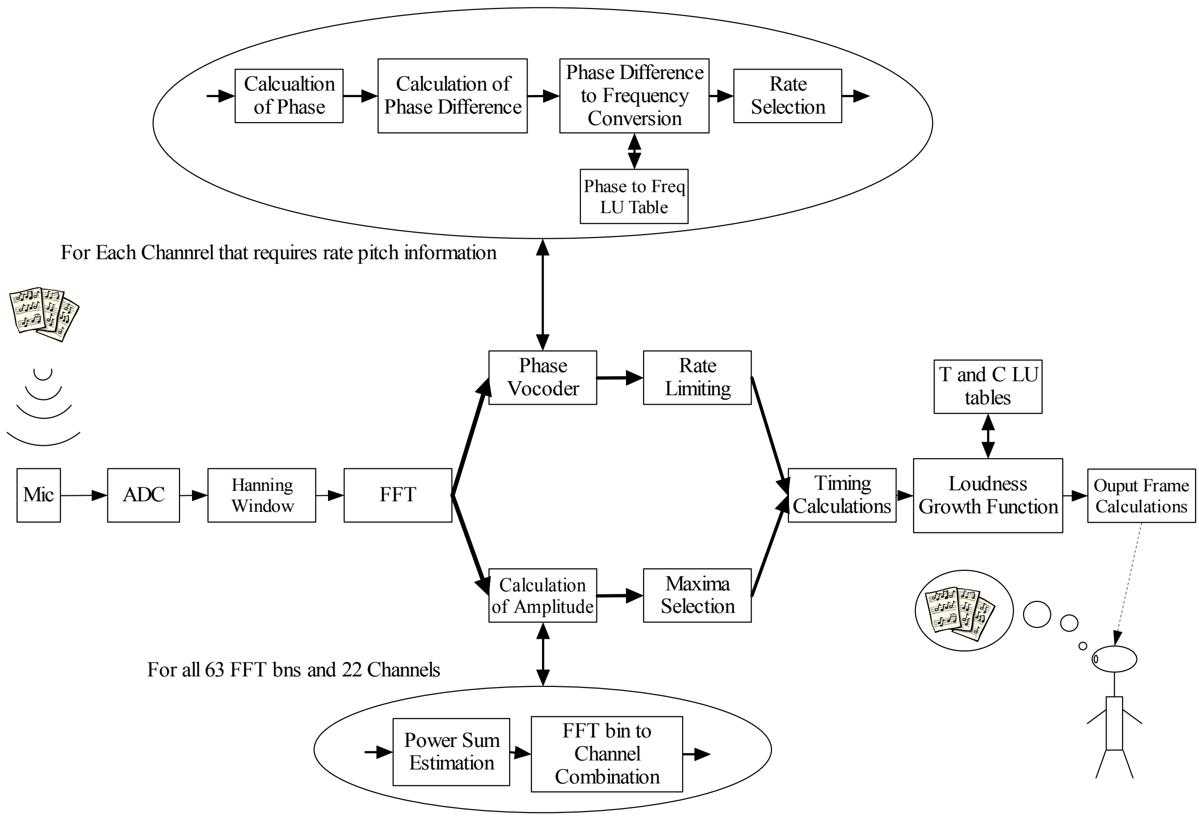


Figure 9-3 System diagram of Music-L signal processing scheme

9.3.1 Microphone input

The microphone used is a standard HS-8 Nucleus microphone which includes a Knowles EL7189 directional element. It has a built in pre-emphasis with a + 6dB/octave up to

~4KHz and -18dB/octave after this. This pre-emphasis is used to reduce signals in the spectrum that are probably not related to speech and at a first order may compensate for the filtering normally provided by the auditory canal.

9.3.2 ADC

The ADC is 16 bit using Sigma-Delta conversion with a programmable sampling rate.

9.3.3 Von Hann window

A window is required to reduce discontinuities in the blocks that are processed by forcing the extremities to zero. The frequency response of the Von Hann window, a raised cosine, has an equivalent bandwidth of 3dB at 1.5 bins. When the sampling rate is 16 KHz and the block size is 128 points the bandwidth of each bin is 125Hz. Thus the effective bandwidth using the Von Hann window is 187.5Hz (1.5bins x 125Hz).

9.3.4 Filterbank

Initially it was decided to use Infinite Impulse Response (IIR) filters that would operate at that sampling rate of 16 000Hz. When this was implemented it was found that the processing power required to perform the filtering and frequency extraction (discussed in 9.3.5) required a processing speed of around 50MIPS. This was within capabilities of the SPEAR but, however, resulted in a battery life of only 2hrs. This may have been useful for short tests but to provide take-home experience, longer battery life was needed, requiring a reduction in the processor speed. Another important reason to not use the IIR filterbank was that to evaluate Music-L it would be required to compare it to a current speech-processing scheme such as ACE. If improvements were found it would not be possible to isolate which part of the processing scheme was responsible. For ease of comparison with the ACE strategy, it decided that as many parameters as possible would be kept identical to focus on the different coding technique used. Thus, a standard 128-point FFT filterbank was used that had an analysis rate of 1000Hz. Using the SPEAR device and the programming software, a comparable ACE strategy could be programmed so that the only differences, were parts of the new Music-L strategy. The 22 channel FFT used in the Voc-L and ACE strategy is shown in Figure 9.4

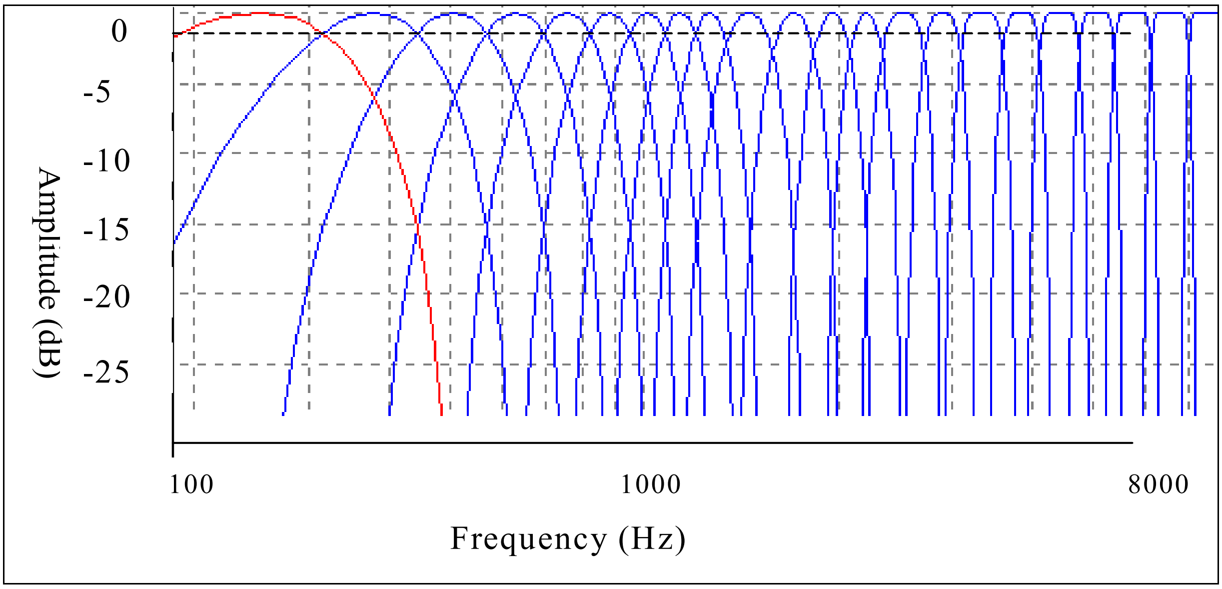


Figure 9-4 FFT Filterbank used in the Music-L strategy

9.3.5 Extraction of multiple frequencies

Zero crossing

The zero crossing technique was initially chosen to extract the frequency from each channel that had an upper frequency limit below 1000Hz. 1000Hz was chosen as the upper limit as a result of Chapter 3 which shows that the pitch has leveled off at this frequency. From each filtered channel, each sample was analysed for its sign and a flag was raised when the sign changed from negative to positive or visa versa. The number of samples was counted from the time of the last sign change, doubled to take account of the half period counted, divided by the sampling rate and inverted which resulted in the extracted frequency in Hz. This scheme proved reasonably accurate especially for the low frequency, narrow band filters which attenuated other frequencies not within that band. To further reduce zero crossings that were not related to the dominant frequency in each band, maximum and minimum zero crossing limits were applied to each band. If a period was counted that was outside the range defined by these limits the period was discarded and counting restarted from zero.

FFT output

When the IIR filtering system was changed to the block processing FFT system the zero crossing counting technique had to be changed. The filtering of the FFT is performed at a much slower analysis rate (1000Hz in this case) where a 128-point block is analysed. The block moves in relation to the incoming audio samples by 16 samples each analysis (Sampling Rate /Analysis Rate). There is an overlap of 112 samples or 87.5% each FFT analysis.

The output from the FFT is a set of complex values that relate to the amplitude and phase of the frequency in each bin of the FFT. The FFT produces 128 bins of which half contain mirrored information, leaving 64 useful bins. Each bin has a bandwidth of 125Hz (Fs/Block Size =16 000/128). The first bin is called DC and has a bandwidth of half this, from 0 to 62.5Hz, which is discarded due to the low frequency. This leaves 63 bins of useful information.

FFT magnitudes and channel combination

The 63 bins of amplitude data are combined into the 22 channels relating to the 22 electrodes. This combination is described by the patient's map which defines which frequency bins will be combined to which channels

The magnitude can be calculated from the $x+iy$ complex components when combining n FFT bins to channels by performing either a:

$$\text{Power Sum} = \sqrt{x_1{}^2 + x_2{}^2 + \ldots + x_n{}^2 + y_1{}^2 + y_2{}^2 + \ldots + y_n{}^2}$$

Or

$$\text{Vector Sum} = \sqrt{(x_1 + y_1)^2 + (x_2 + y_2)^2 + \ldots. + (x_n + y_n)^2}$$

In this implementation a power sum was used to keep the process the same as that normally implemented with the ACE strategy.

Phase vocoder

The output from each FFT bin also contains phase information about the frequency within each bin that is normally discarded by the ACE strategy. This phase information can used to extract frequency information by a system known as a phase vocoder. A block diagram of the implemented phase vocoder is shown in Figure 9-6. The Voc-L strategy has 5 channels below 1000Hz and uses bin outputs 2-6 in a 1:1 relation. The Instrument-L strategy, however, requires 10 channels below 1000Hz. By reducing the sampling rate to 12KHz, the FFT bin bandwidth is reduced from 125Hz to 93.75Hz. This provides 7 FFT bins in the desired frequency range ~150-800Hz. For further division of frequency the second, third and fourth bins were each split into 2 ranges of rates extracted from each bin as shown in Table 9.2. The range of rates extracted from these split channels is overlapped to provide the desired range of rates on each channel. This increased the effective number of channels to 10. This overlapping of low frequency channels may also provide more information relating to the pitch of the incoming signal by presenting frequency information to more than one channel.

Channel	FFT Bin No	Bin low Freq (Hz)	Bin High Freq (Hz)	Rate extracted (pps)
1	2	140.675	234.375	150 –243
2				178-271
3	3	234.375	328.125	212-305
4				253-346
5	4	328.125	421.875	300-393
6				356-449
7	5	421.875	515.625	424-517
8	6	515.625	609.375	504-597
9	7	609.375	703.875	600-693
10	8	703.875	791.875	713-848

Table 9.2 Frequency allocation and rates extracted from each bin of the variable rate channels for the Instrument-L strategy

To visualize how the phase vocoder works, the vector, x_1+iy_1, produced by the output of a bin after an FFT is performed, is shown in Figure 9-5. The phase of this vector is $\theta_1 = \arctan(y_1/x_1)$. The next time the vector is calculated by the FFT, the block has moved on 16 samples and the vector has moved to x_2+iy_2. The phase of this angle is $\theta_2 = \arctan(y_2/x_2)$. The angle difference between these two vectors is $\Delta\theta = \theta_1 - \theta_2$. The rate of rotation of this vector, $\Delta\theta$ /time, is related to the frequency within that bin by the formula

Frequency = $\dfrac{\Delta\theta.Fs}{2.\pi.Hop}$, where Fs= Sampling Rate and

Hop= Fs/Analysis Rate

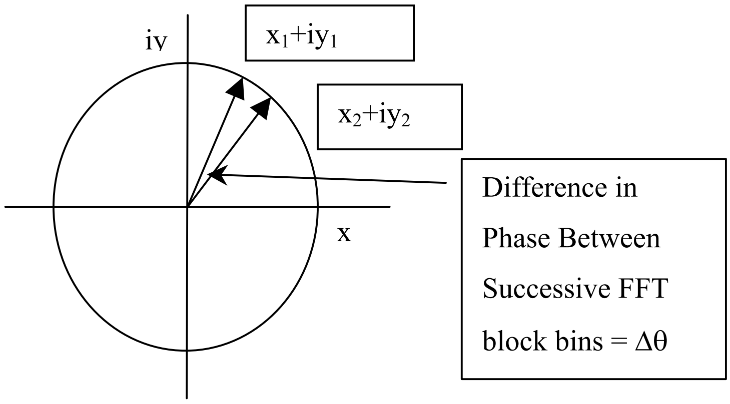


Figure 9-5 Vector plots showing change in phase of successive FFT operations.

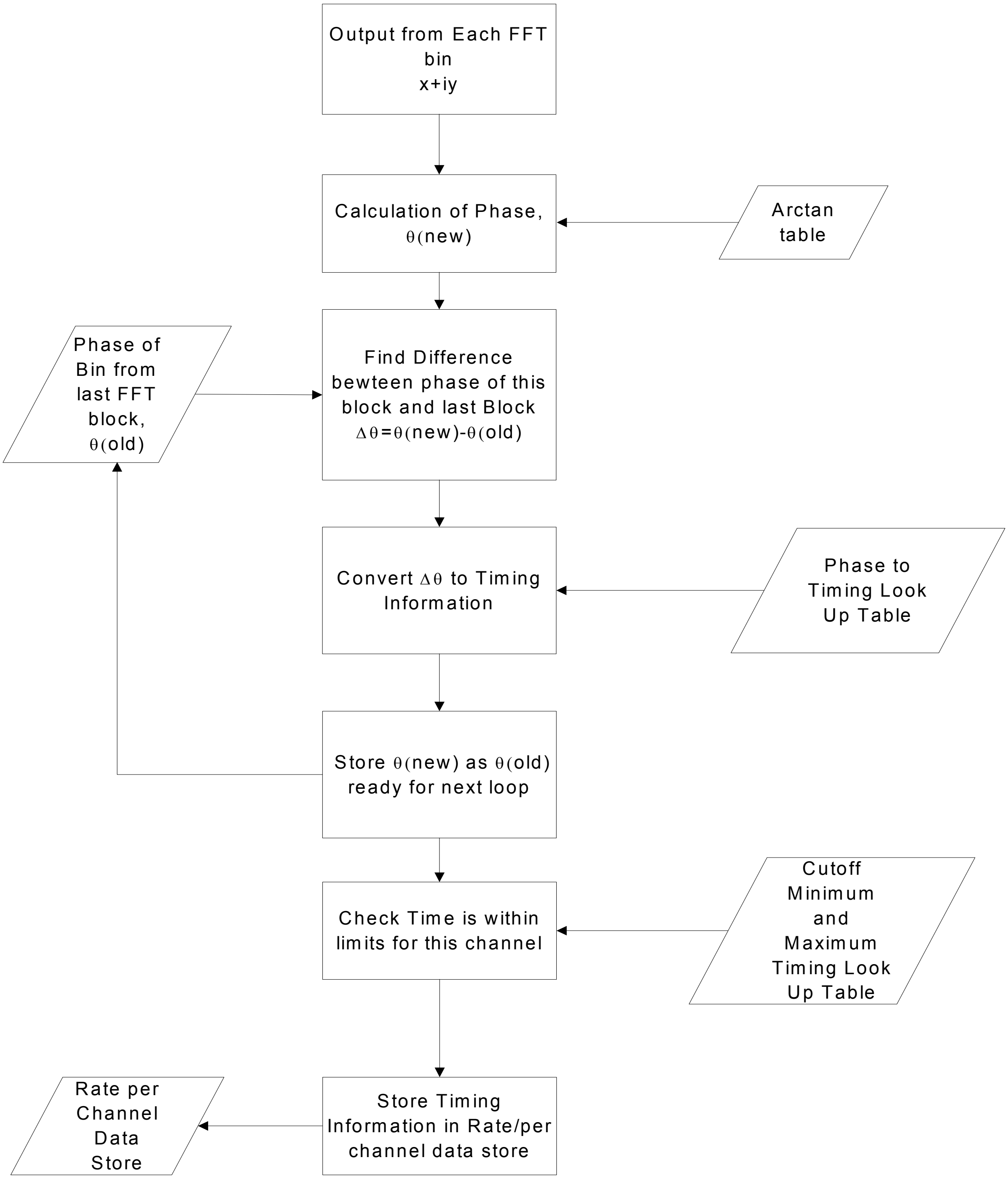


Figure 9-6 Block diagram of implementation phase vocoder

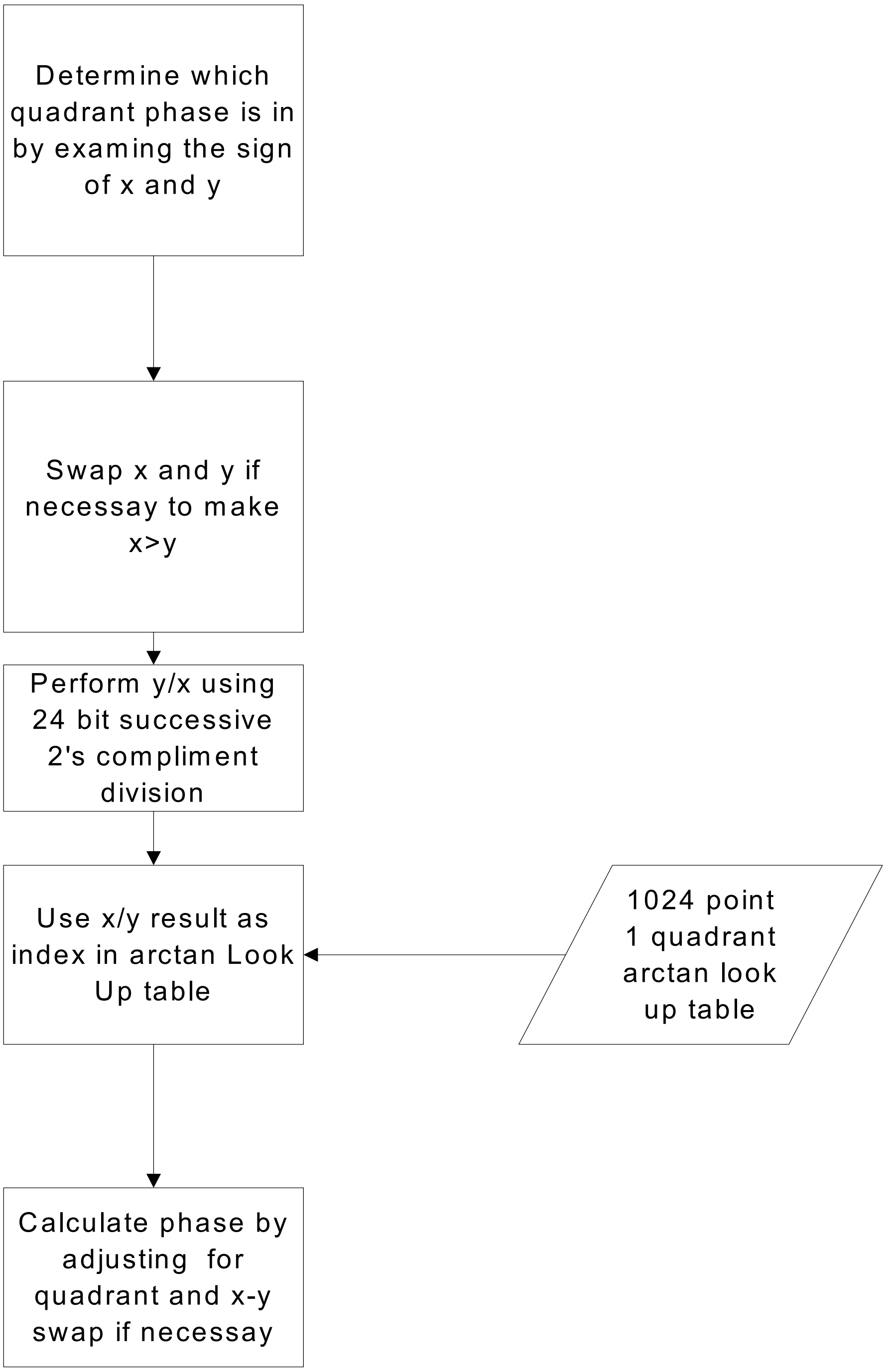


Figure 9-7 The 'Calculation of Phase' block from Figure 9-6 in detail

If the input is a 625Hz sine wave (center frequency of bin 6) the extracted frequency using the phase vocoder is shown in Figure 9-8.

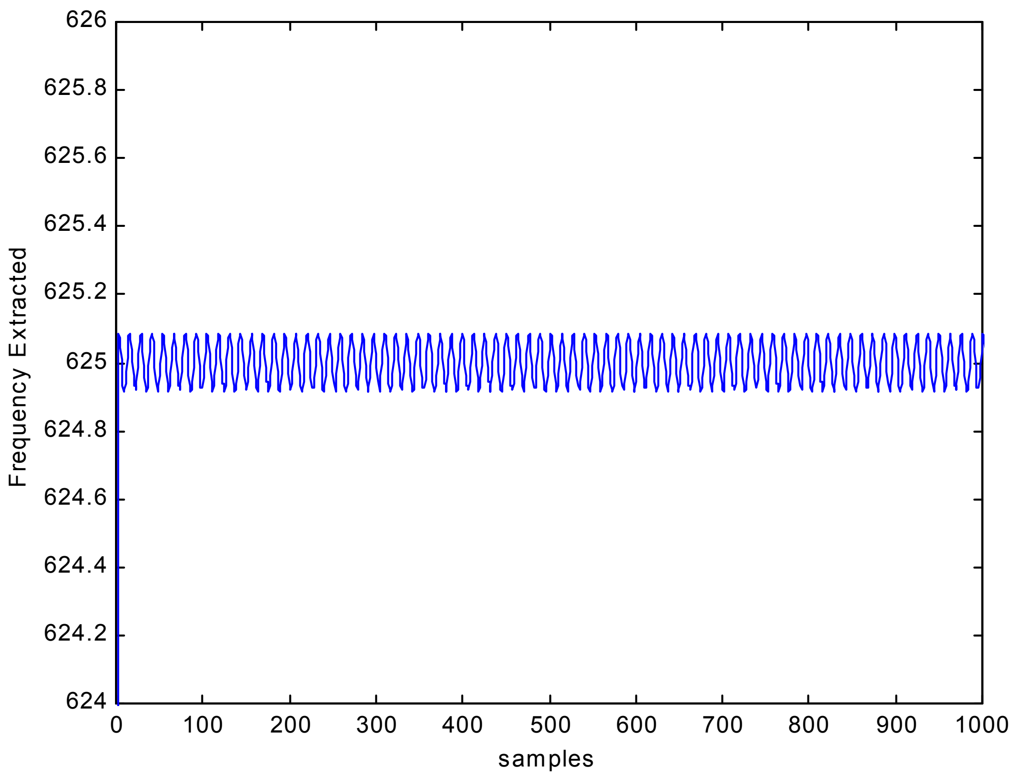


Figure 9-8 Phase vocoder output in response to a 625Hz (center frequency of bin 6) sine wave. The maximum error is of the order of 0.1Hz.

When a sine wave is swept through the region of 500Hz to 750Hz (½ bin below and ½ bin above bin 6) the plot shown in Figure 9-9 reveals how linear and accurate the phase vocoder over this range.

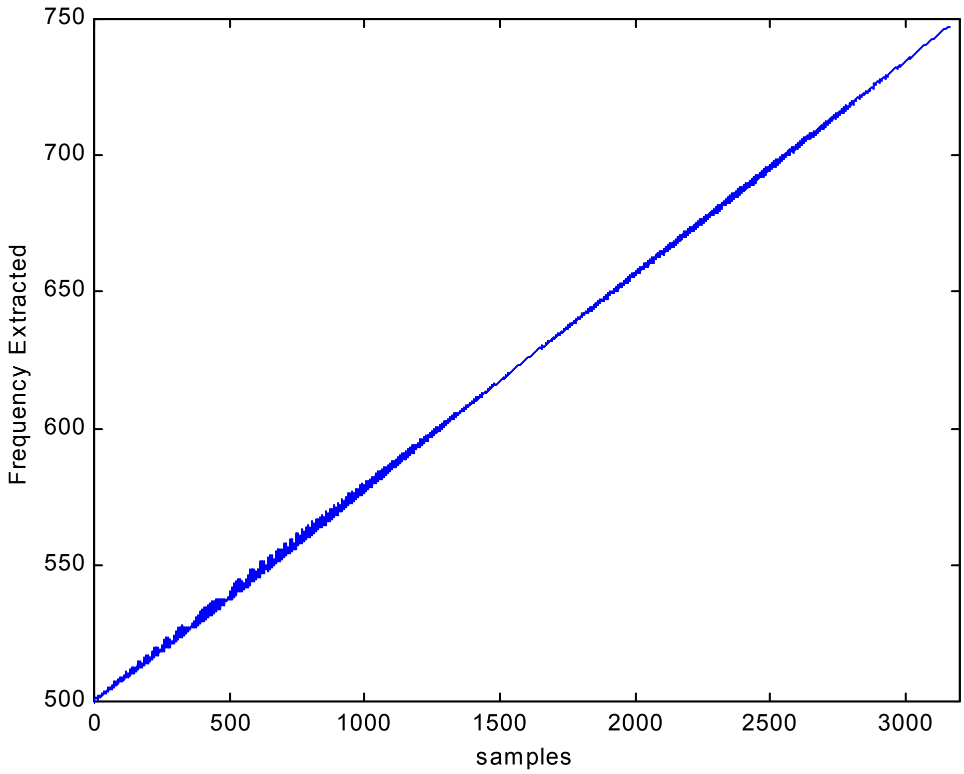


Figure 9-9 Output of phase vocoder as frequency is swept between 500Hz and 750Hz

The frequency extracted by the phase vocoder is converted to a period with a resolution of 0.2μs (the resolution of the implant). For example a 200Hz frequency would be represented by the period 25 000. This period extracted by the phase vocoder for each channel is placed in a buffer called Stim_Time. If the stimulation rate is 1000pps the period of each output frame when expressed in units of 0.2us is 5000.

9.3.6 Maxima selection

The amplitudes of all the channels are scanned and the largest 8 maxima are selected.

9.3.7 Timing information

Each variable rate channel has a buffer called Current_Time storing its current value of the time when it should stimulate to maintain the current period stored in the buffer Stim_Time. The buffer Stim_Time is updated by the phase vocoder once each analysis period. The Current_Time buffer, however, is updated once a channel has stimulated. The new Current_Time is then calculated based on the time when it stimulated and the Stim_Time period calculated by the phase vocoder. This is when any changes in frequency will be updated. After the maxima selection routine, the selected channels are then scanned to check whether the Current_Time this channel would stimulate is within the frame time.

An example the output of the maxima selection routine for two selected variable rate channels may be:

Channel	Current_Time	Stim_Time	Frequency extracted
22	9000	20 000	250
18	2000	7143	700

Table 9.3 Example of channels being selected based on their timing information.

For the next frame, channel 18 would be stimulated, as its Current_Time (2000) is less than the frame time of 5000. The result would be channel 18 would be stimulated at the time point 2000 into the frame. Channel 22's Current_Time (9000) is outside the frame time (5000) and not selected for stimulation. This channel however has its Current_time updated by subtracting the frame time of 5000 from the Current_Time resulting in a new Current_Time of 4000. Channel 18's Current_Time was updated differently because it was stimulated. The Current_Time has the Phase Vocoder output of 7143 (700Hz) added but also the last time position subtracted i.e. Current_Time(new)=Stim_Time-(frame_time-Current_Time(old)), in this case Current_Time(new)=7143-(5000-3000)=5143.

Channel	Current_Time	Stim_Time	Frequency extracted
22	4000	20 000	250
18	5143	7143	700

Table 9.4 Updated Current_Time for example in Table 9.3 ready for the next frame of stimulation.

In this next frame both Current_Times are less than the frame time of 5000 so both channels will be stimulated.

The following flow diagram shows how the timing information is calculated after the maxima have been selected.

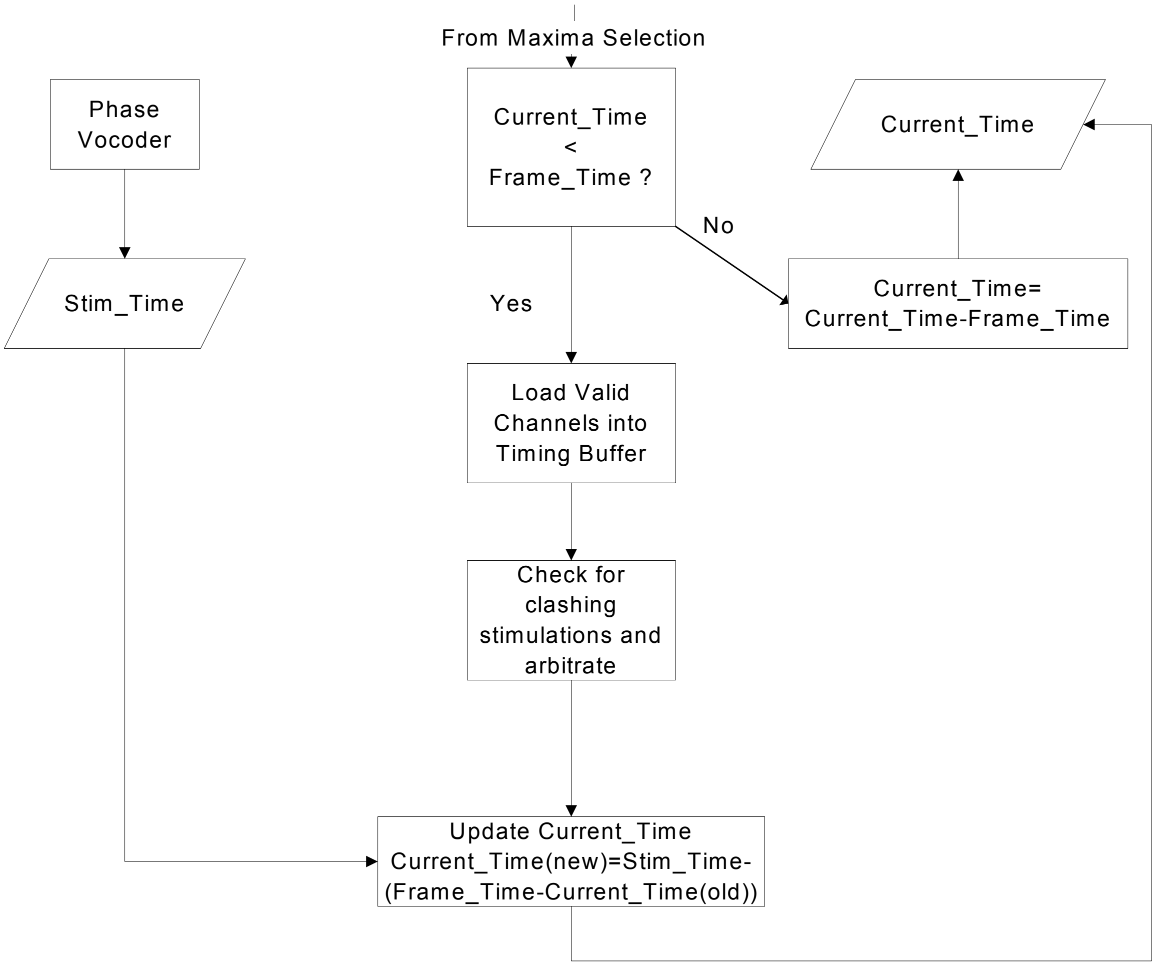


Figure 9-10 Flow diagram of how timing is calculated

Each stimulation pulse has a length of 65µs or in unit of 0.2µs, 325. If two stimulation pulses overlap in the requested time, arbitration must be carried out to decide which stimulation to delay or drop. In practice it was found that drops or delays in stimulation pulses had to be rarely performed.

The arbitration followed some general rules.

If the channels overlapping consisted of a low frequency, variable rate channel and a high frequency constant rate channel, the variable rate channel always took priority with the high frequency channel being delayed. The low variable rate channels are designed to provide pitch information and thus are given priority over high frequency channels that are providing less pitch and perhaps more timbral information.

If the channels overlapping consisted of a two low frequency variable rate channels the lower frequency variable rate channel has priority. This is a direct result that low frequency stimulation is more sensitive to jitter as shown in Chapter 4.

If the channel delayed was a variable rate channel and the delay was more than 8% of the period, this stimulation was dropped. 8% was the minimum jitter discriminable amount found from experiments from Chapter 4.

If the channels overlapping consisted of a two high frequency, constant rate channels, the channel with the largest amplitude had priority.

9.3.8 Amplitude to current level conversion

To convert the amplitude to a current level to a linear increase in loudness, a function called a loudness growth function is applied to the amplitude of each channel. This function is defined by three parameters:

i. b, base level set at 4.

ii. m, level at which the amplitude growth function reaches a ceiling, set at 150.

iii. Q factor, defined by the point on the loudness growth function where a Q% drop in level is defined by $m/\sqrt{10}$.

This is a standard function applied to all signal processing programs such as ACE and SPEAK. As shown in Figure 9-11, for a range of values of b, m and Q, this gives a quasi-logarithmic compression.

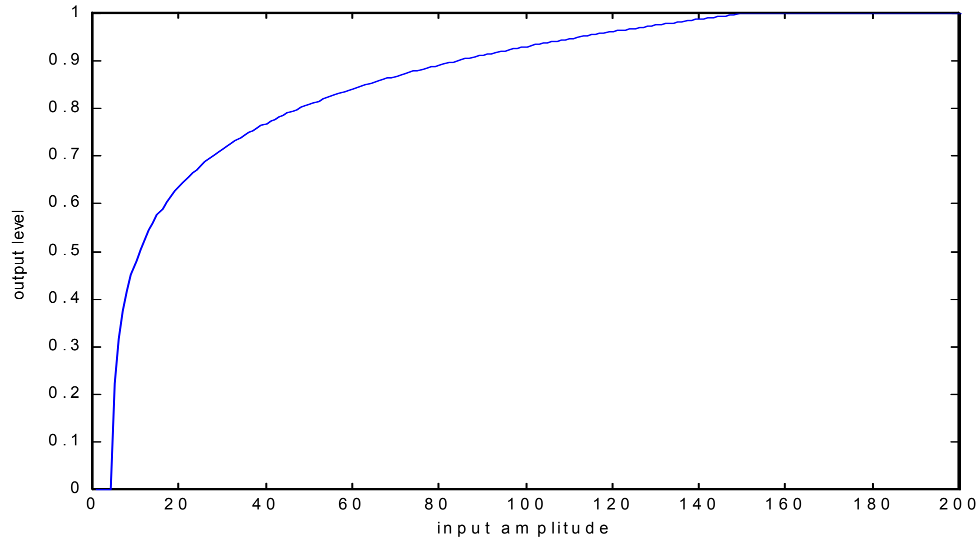


Figure 9-11 Loudness growth curve mapping input amplitude to output current level

9.3.9 Calculation of Output Frames and Electrode Stimulation

9.4 T and C evaluation

A scheme was implemented which adjusted the amplitude to allow for changes in loudness due to changes in the rate of stimulation. Using monopolar stimulation, the T and C levels follow a reasonably smooth decrease in level as rate increases, although this is individual for each subject (For a explanation of T and C levels see section 1.5 Electrical stimulation of the cochlea in Chapter 1). When a CIR is mapped, T and C levels are measured at a particular rate (the rate of constant stimulation). In the case of Music-L the rate is variable and hence the loudness will also change as the rate of stimulation tracks the incoming frequency. To account for this, T and C levels could be measured for each possible stimulation rate allowable on each channel (infinite), or T and C levels could be measured and an approximate function that predicts how the levels change as a function of rate could be applied. The latter solution was implemented however; it was found that the T and C levels do not follow the same function. The C level usually reduces as the rate increases but the T level reduces by an even larger amount. To account for this, a level was chosen at 80% of the way between T and C. Measurements were taken from CIR's for 3 different rates for all variable rate channels. These rates consisted of the lowest, middle and highest rate allowable on each channel. For example, channel 4 can stimulate between 250 and 500pps, therefore measurements were taken at 250, 375 and 500pps. The 80% point was calculated and a first order polynomial was applied to the function to smooth it. This function was then used to modify the loudness as the rate changed on this channel. The function was implemented as a set of look-up (LU) tables with one table required for each channel that has a variable rate.

This scheme was compared with a scheme that did not make allowance for changes in rate but where T and C levels were gathered at a rate equal to the center frequency of that channel. A CIR was exposed to both strategies by playing swept sine waves and short pieces of music and speech. The CIR reported no noticeable differences between the two schemes.

It was decided not to keep the amplitude function scheme, as the subject appeared to receive no benefit from it. It is noted that the T and C levels varied by only around 5-10% over the range tested and the lack of benefit was thought due to the small change applied by the amplitude reduction scheme. CIR's who use bipolar stimulation often have T and C rate functions which can vary considerably more than when using monopolar mode. It may be that a CIR using bipolar stimulation may received more benefit from the amplitude reduction schemed described above.

9.5 RF capture and Verification

To verify the operation of the strategy the processor was subjected to various stimuli and the stimulation pulses that would be normally applied to the CIR were captured and displayed as an electrodogram.

Two sine waves at 262Hz and 600Hz were applied to the processor and the output captured. From the electrodogram shown below the two different rates on the two groups of electrodes can be clearly seen. The lower frequency electrodes, 21, 20 and 19 have the low frequency 262Hz rate stimulation while the higher frequency 600Hz is stimulated at 600pps on electrodes 14 and 15.

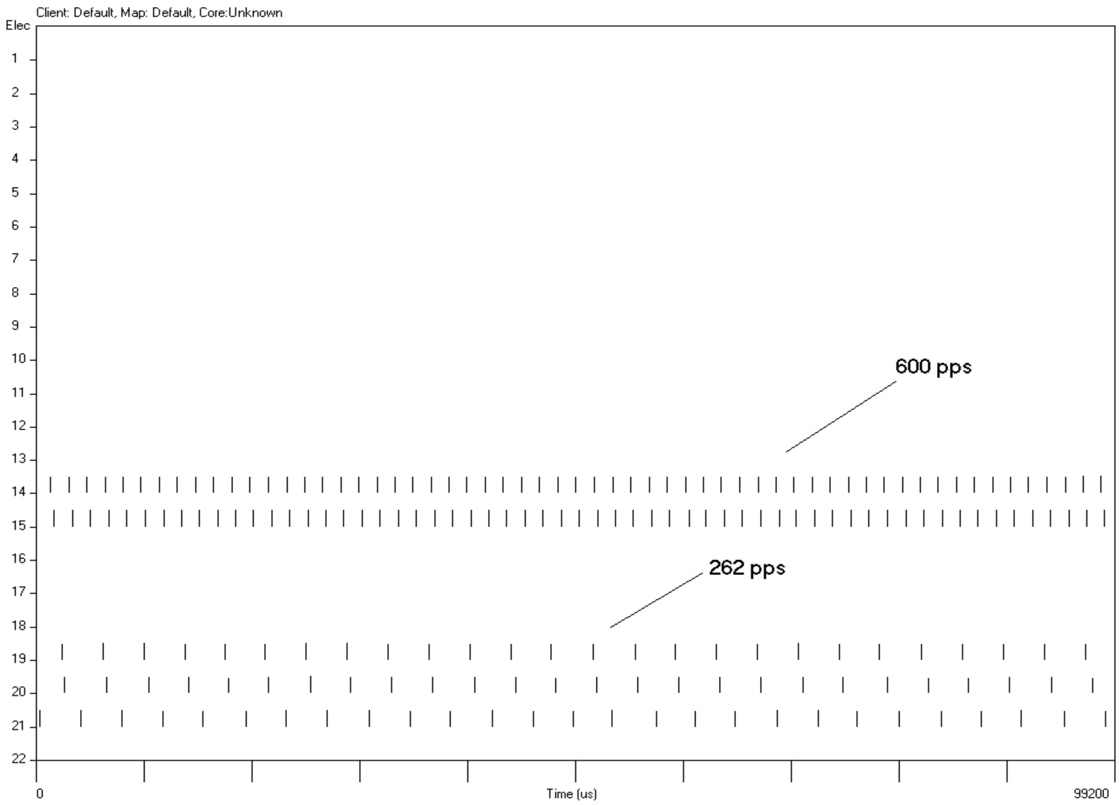


Figure 9-12 Sine wave input of 262 Hz and 600Hz (x tick is ~0.01s).

A sine wave was swept from 100 to 1000Hz while observing the output for the 5 variable rate channel strategy Voc-L. The rate produced on each electrode is the frequency extracted from the sine wave. The electrodogram also shows the overlap of each filter.

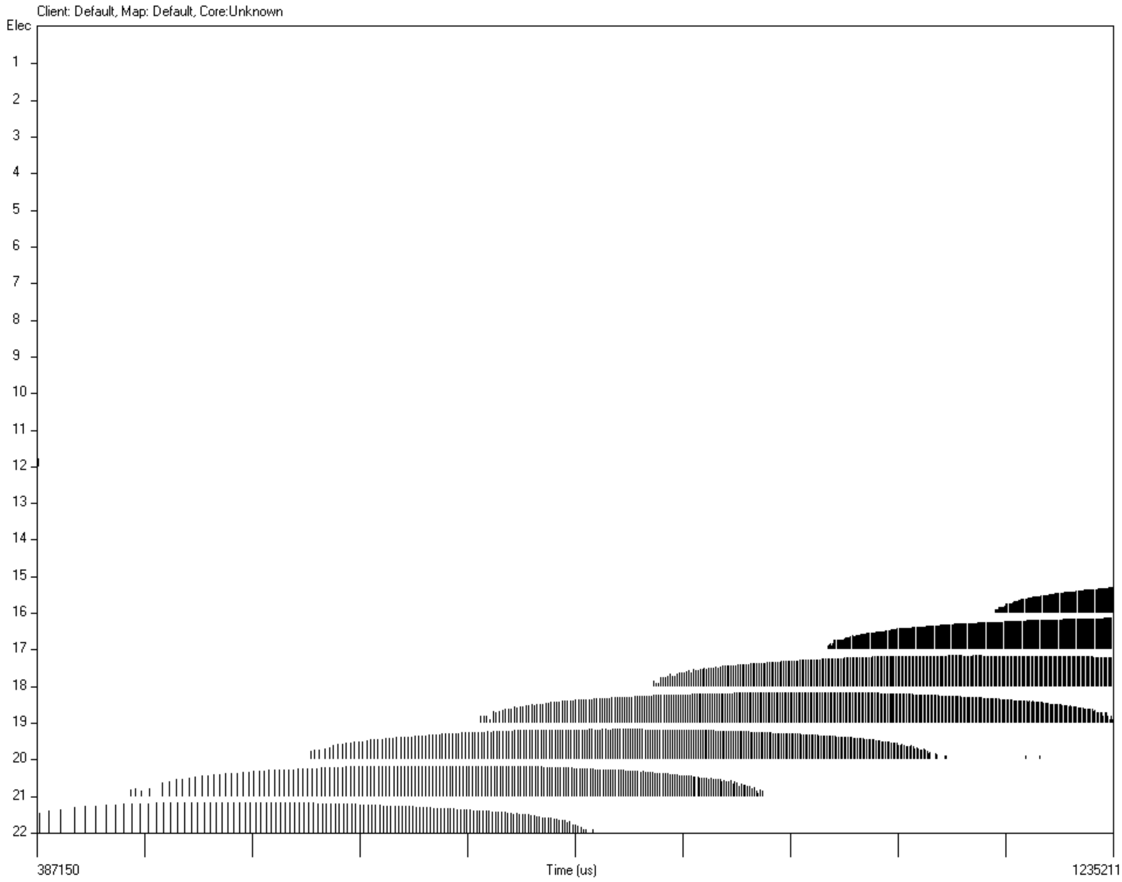


Figure 9-13 Sine wave input sweep 100Hz to 100Hz for the 5-channel Voc-L strategy (x tick is ~0.085s).

The statistics gathered for the sweep are shown below.

Channel	Min Freq pps	Max Freq pps	Average Freq pps
22	103	380	250
21	190	508	406
20	242	637	522
19	277	763	646
18	577	812	711
17	1013	1016	1014

Table 9.5 Stimulate rate statistics gathered from the sweep shown in Figure 9.13.

Two sine waves that were close in frequency (200 and 240Hz) were combined and the result is shown in Figure 9-14.

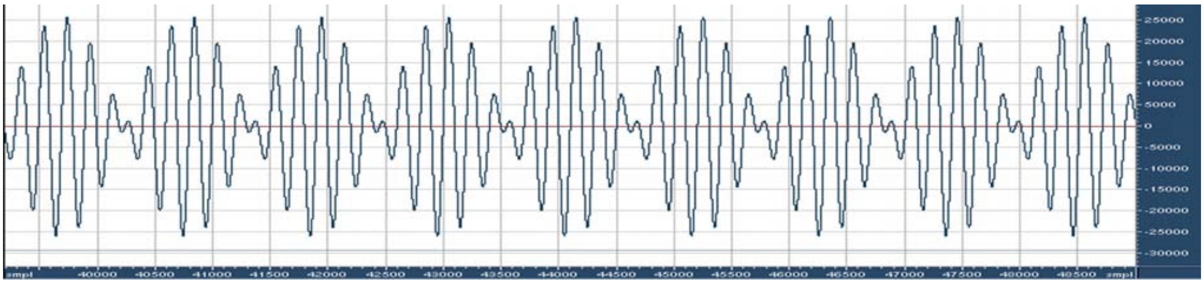


Figure 9-14 Combined 200 and 240Hz sine waves.

This signal was applied to the strategy and the electrodogram is shown in Figure 9-15.

It can be clearly seen the rate of stimulation of the low frequency channels is the 200Hz lower frequency while the rate is amplitude modulated at the difference frequency 40Hz (240Hz-200Hz).

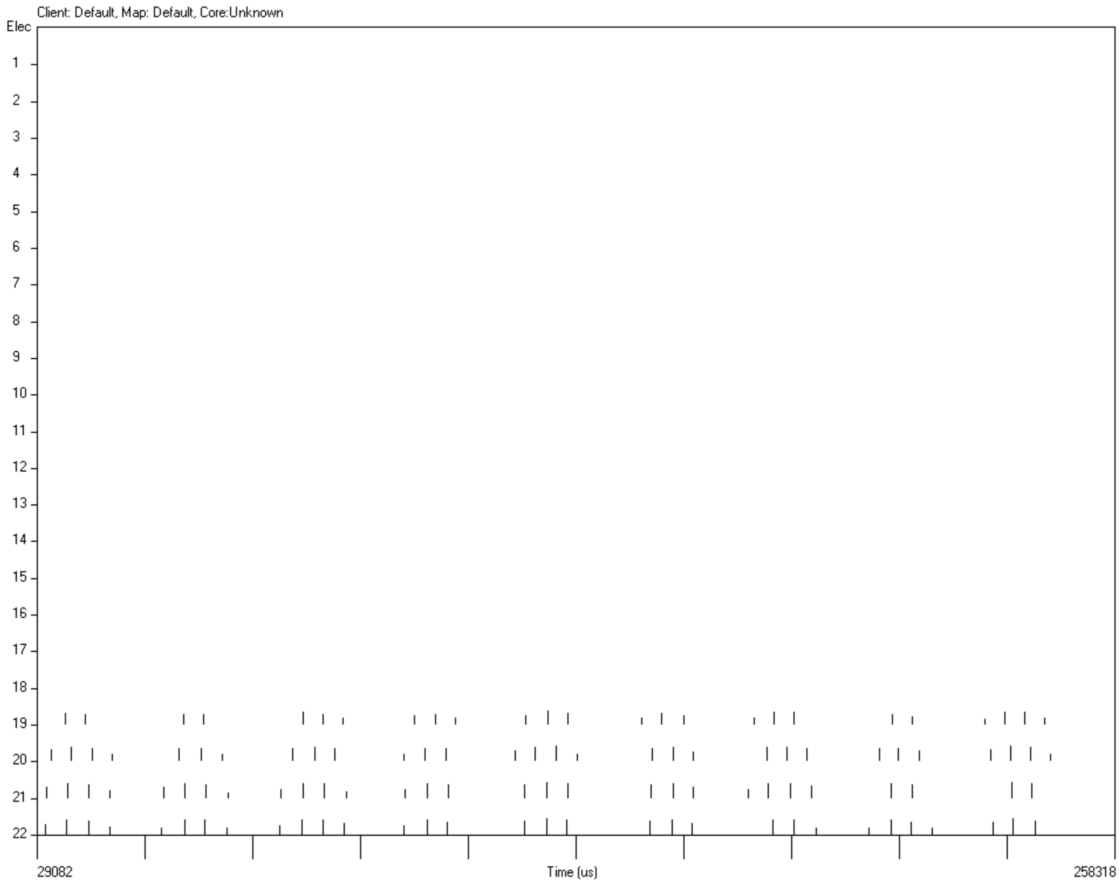


Figure 9-15 2 Sine waves, close in frequency 200 Hz and 240 Hz. The stimulation rate is 200Hz and is amplitude modulated at the difference frequency of 40Hz (x tick is ~0.02s).

Sampled real instruments, with the pitch of middle C (262Hz) as played by a piano, trumpet and violin were applied to the strategy. The electrodogram is shown in Figure 9-16. The differences in electrodes stimulated for this same pitch vary markedly. The sustained part of the piano has a strong fundamental and some transient high frequencies. The trumpet has a very weak fundamental and is dominated by stronger higher harmonics typically associated with brass instruments. The violin also appears to have stronger higher harmonics than the piano.

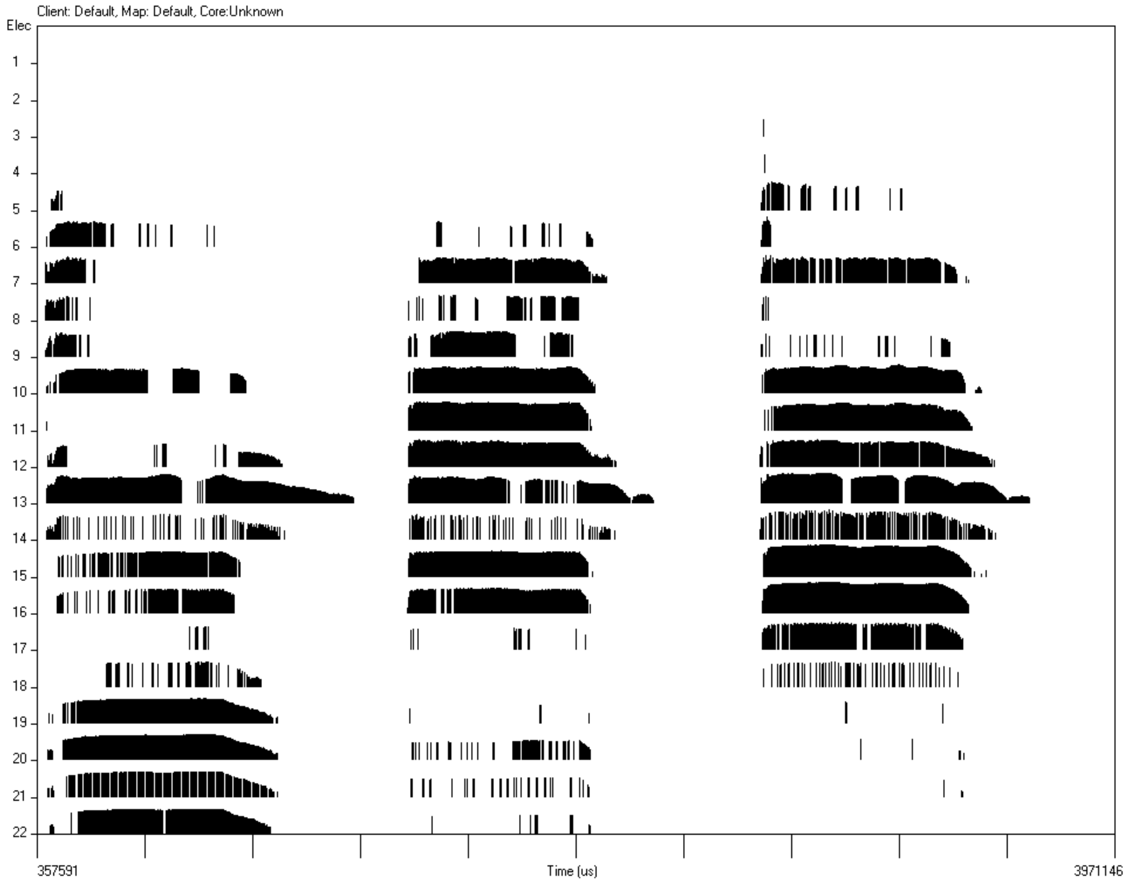


Figure 9-16 Middle C as played by piano, trumpet and violin (x tick is ~0.4s)

9.6 Discussion

This chapter described a new signal processing strategy, named Music-L that was developed to improve music perception. The strategy implements some of the results gained from previous chapters of this thesis.

Chapter 3 'Temporal Pitch- Place Pitch' helped determine the appropriate electrodes for rate stimulation to achieve changes in the pitch percept and to find the best sounding (quality) position for rate stimulation.

Chapter 4 'Jitter Discrimination' helped determine the limits to which pulses could be delayed or dropped.

Chapter 7 'Pitch Discrimination of Pure and Complex Tones' helped determine the improvement necessary that this new strategy must provide.

The major difference between the Music-L and current speech processing strategies like ACE is in the way stimulation is delivered. In a strategy like ACE, the stimulation on each channel is at a constant rate. In the Music-L strategy, however, the rate of stimulation of several low frequency channels is determined by the frequency extracted from their respective channel filter.

It is intended that the stimulation rate combined with the electrode position will elicit a stronger and more accurate perception of pitch than is currently done by just using a strategy like ACE that relies solely on the place pitch principle.

A short trial to determine benefits gained by the Music-L strategy follows.

Chapter 10
Music-L Pilot Studies

10.1 Introduction

The Music-L strategies, Voc-L and Instrument-L, were programmed with volunteer recipients and some preliminary tests performed. A full test of these strategies is beyond the scope of this thesis as it may take up to one year to complete an A-B-B-A type comparison that includes comprehensive training and test materials to identify any benefits of these new strategies. The type of training material required to be developed is outside the discipline of this thesis and requires more experience in musical education. One software training package is known to exist (Gfeller et al., 2000) and it is hoped that this may be used at a later date to further evaluate these strategies. The type of tests that can be done is qualitative evaluations and pitch discrimination tests previously described. The results of these tests may provide some information as to whether the extra rate information is utilised and whether the sound quality has improved.

10.2 Pilot study 1

10.2.1 Three octave Pitch Discrimination Test comparing ACE and Instrument-L

The three octave PDT can be used to monitor any overall changes in pitch discrimination. This test, described in chapter 8, was used with one subject over the course of two months. In the second month of this period the subject took the Instrument-L strategy home. An audio cassette was prepared for this subject consisting of MIDI songs that the subject had previously identified on a questionnaire as familiar and likeable. These songs contained no voice. If a vocal melody was present this was replaced by an instrument. Voice was not included as it is has been shown that the comprehension of the sung speech is used to recognise the music as apposed to the melody itself and hence avoids a potential confounding variable. The subject listened to the tape for around ½ an hour, 3-4 evening per week. Over the period of 8 weeks the subject visited the testing center and undertook the three-octave PDT using the Instrument-L strategy. Every second week the three octave PDT was undertaken with both the Instrument-L and the subject's normal ACE strategy. The piano stimulus was used.

10.2.2 Results

The result of the three octave PDT is shown in Figure 10-1. The subject initially scored around nine semitones with virtually no exposure to the new strategy. This PDT score measured, rose to around 16 semitones during test session in week three, and then began to improve (reduce), achieving 4-5 semitones in the final weeks of the trial. The initial test with the ACE strategy revealed a PDT score of around an octave. Over successive weeks this score remained around the same using.

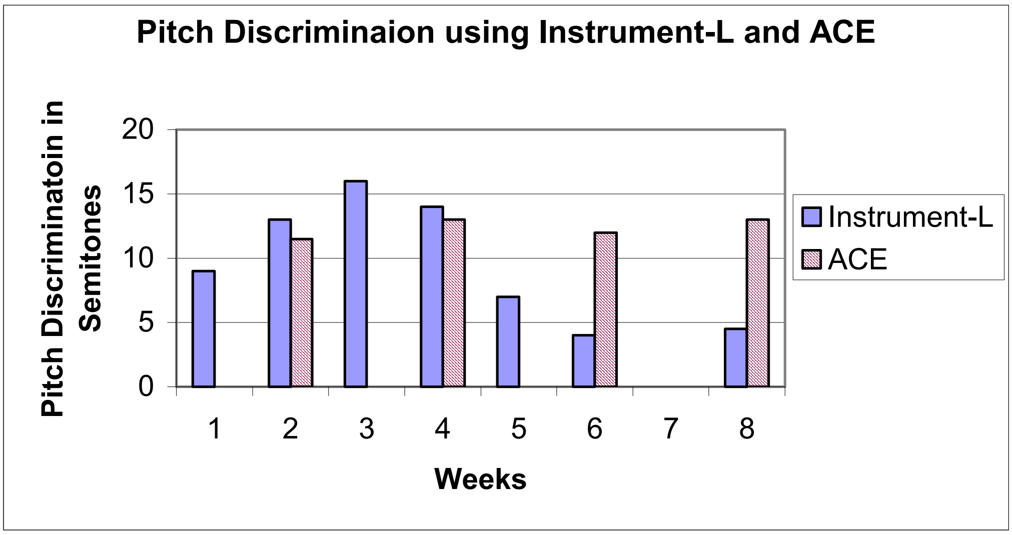


Figure 10-1 Results of three octave PDT with one participant using the ACE and Instrument-L

10.3 Pilot Study 2

10.3.1 Isolated frequency PDT and JND test comparing ACE and Voc-L

One subject was tested with the isolated frequency PDT and JND test, while using the Voc-L and ACE strategies. The difference between these strategies, discussed in detail in chapter 9, is the ACE strategy uses fixed rate stimulation on each channel whereas Voc-L used variable rate stimulation on the 5 lowest frequency channels that are derived from the frequencies in the signal.

10.3.2 Results

The results of Same/Difference JND test at 400Hz are shown in Figure 10-2. It appears that the subject was able to discriminate better (lower %) using the Voc-L strategy than the normal ACE strategy. The level achieved using the Voc-L strategy (0.41%) was repeatable the next week (0.47%).

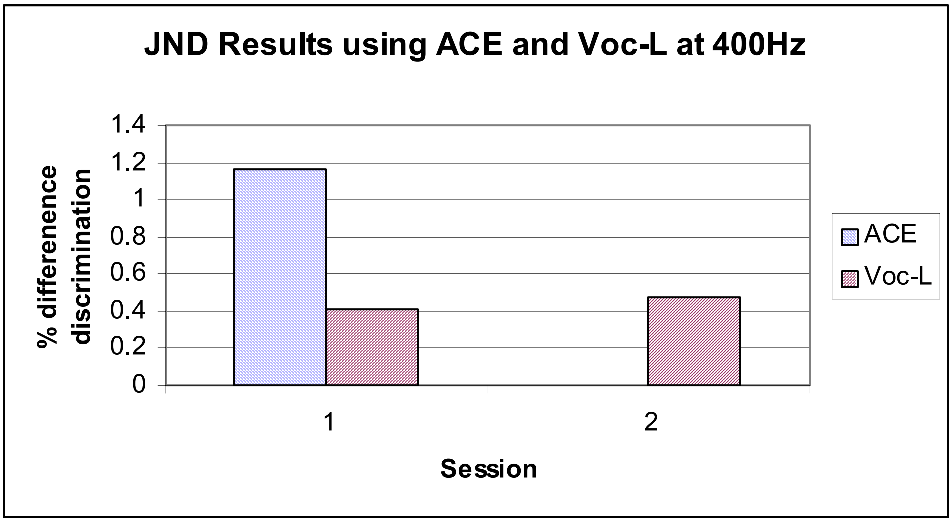


Figure 10-2 Results of JND test at 400Hz using ACE and Voc-L

The results of Higher/Lower PDT test at 400Hz are shown in Figure 10-3. Upon initial exposure it appears that the subject can identify pitch direction better using their normal ACE strategy (0.29%) compared to around a semitone (6.1%) using the Voc-L strategy. As with the first subject tested, discussed earlier, this discrimination ability appeared to improve (decrease) over successive test sessions (6.1% to 3.1% to 1.4%). This subject had no take-home experience and only had exposure to the Voc-L strategy during the 1-2 hour test session. The level achieved using the Voc-L strategy appeared to improve but not yet to the level available with their normal ACE strategy. This subject was not available to take part in further testing due to other commitments.

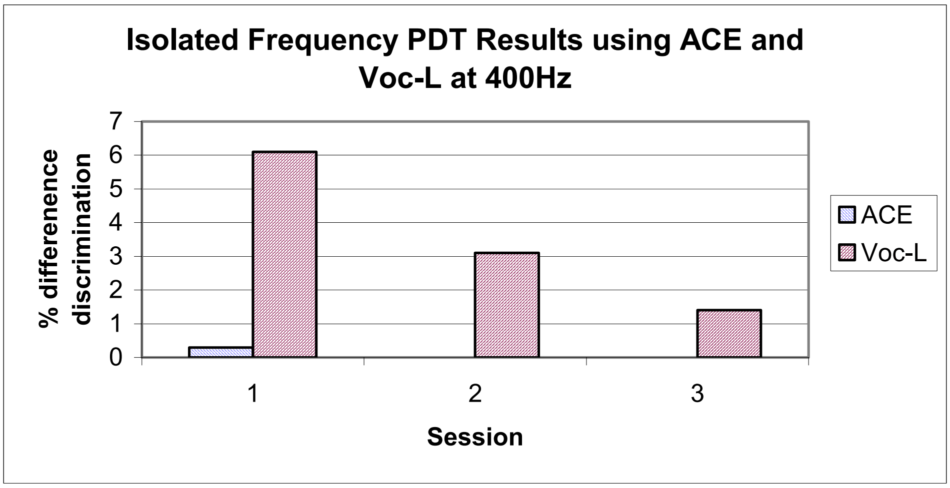


Figure 10-3 PDT at 400Hz using ACE and Voc-L

10.3.3 Discussion

It appears that the Instrument-L strategy appeared to provide benefit to the first subject tested by improving their pitch discrimination ability over the 8 week period. It is of interest to note that it took almost one month before any improvement was seen. This may be due to the subject not being exposed to any rate pitch information since becoming deaf and it may take some to utilise the new information.

With the second subject it was found that the ability to determine if two tones were the same or different was immediately better with the Voc-L strategy than with their normal ACE strategy. This would indicate that the subject could utilise the new rate pitch information to help discriminate between pitches. When required to indicate the direction of pitch between two tones it was found initially the subject could perform better using their normal ACE strategy, (0.29% compared with 6.1%). Upon further testing over successive sessions it was found, however, that the ability to discriminate direction of pitch did improve in just two sessions to 1.4% but this was not yet at or better than the

0.29% achieved with the ACE strategy. It is hypothesised that since the same/difference test was immediately better using the Voc-L strategy and it has been previously found that direction of pitch does take some time to relearn how to utilise the new rate pitch information, that the levels achieved using the Voc-l strategy might improve with more exposure that would be gained with take home use.

10.4 Anecdotal reports.

Four recipients have participated in evaluating the Music-L strategy.

All subjects have reported that both Voc-L and Instrument-L sound better for listening to music than ACE. Speech however, sounds the best using ACE. Three subjects were asked to identify the best strategy for three listening conditions. The unanimous results are shown in the table below. It should be noted that the name Voc-L and Instrument-L were applied after these qualitative results were found.

Condition	ACE	Voc-L	Instrument-L
Speech	✓		
Music with Vocals		✓	
Instrumental Music			✓

Table 10.1 Preferred strategy for different listening conditions

Comments reported using the Music-L strategies include:

i. “Instruments sounded more real”.

ii. “Singing voice appeared to sing the right notes” and

iii. " You guys are going to have to SHOOT me to get this processor (Voc-L) back! I love it!!".

One subject was tested using Consonant-Nucleus-Consonant (CNC) speech tests using Voc-L and ACE. Using ACE this subject scored 70% correct and on first exposure to Voc-L scored 62%. This score, although worse than ACE, is still a very good score especially when the subject had little time to adjust to the new strategy.

10.5 Pitch dimension extension study

Two subjects participated in a pitch scaling experiment similar to those described in Chapter 3. In this experiment the rates used in the Voc-L strategy were combined with rates used in the ACE strategy to observe how the pitch estimates differed for the two sets of rates. The ACE rates were set as 1000pps/chanel whereas the lowest 5 electrodes for the Voc-L rates are set to 775,625,500,375 and 250 pps.

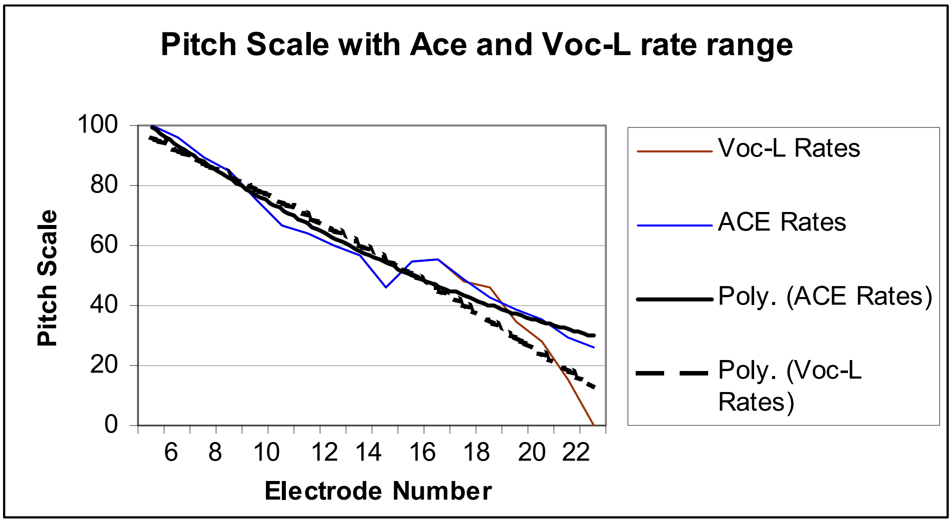


Figure 10-4 Pitch scaling for ACE and Voc-L rates

The data from 2 subjects were scaled, averaged and are shown in Figure 10-4. A polynomial trendline was added to the data to show the difference between pitch estimates for the two ranges of rates. It can be seen that the Voc-L range of rates, as expected, produce lower pitch estimates for the deepest inserted (high numbered) electrodes. The rates used in the Voc-L strategy appear to increase available pitch range by around 17%. If the pitch range is extended it must mean that the pitch on each of these electrodes has increased its perceptual dissimilarity from the next. This may aid pitch discrimination and may also aid electrode discrimination that may lead to improved speech perception.

10.6 Conclusion

This chapter was not intended to be a complete evaluation of the new Music-L strategies as this in itself will be a considerable study. It was however included to provide some initial results from some subjects that have begun to evaluate the new strategies. It does appear that the Music-L does provide improved discrimination scores immediately when required to determine if two tones are the same or different. This may be due to the new rate pitch information increasing the perceptual distance between the two tones allowing easier identification. When subjects are required to indicate the direction of pitch between two tones it appears that it may take some time to be able to associate the new discriminable rate pitch information with a high or lowness. It does appear that subjects can initially detect a difference in the tones due to the different rates but that it takes some time to apply labels of higher and lower to the new rate pitch information. The pitch scaling experiment performed with the rates that are used in the ACE and Voc-L strategies demonstrate how the lower rates on the low frequency electrodes can increase the range of pitch while still providing a continuous relationship i.e. with no pitch reversals due to the different rates found. The anecdotal evidence appears to support that the Music-L strategies do provide some benefit to CIRs with no recipients rating it above ACE for listening to music. Further comprehensive training and evaluation is required to measure the full benefits of the new music strategies.

Chapter 11
Conclusions and Suggestions for Further Work

Rate-Place Pitch

Studies on post-lingually deafened CIRs, using both the most apical electrodes, and also on widely spaced electrodes, show that perceived pitch decreases approximately linearly with insertion distance of the place of stimulation, in agreement with the usually reported tonotopic response of the normal ear. At low stimulation rates, perceived pitch also increases approximately proportional to the log of the stimulation rate. The dependence of pitch on rate saturates towards 1 KHz. The quality of the percept is maximal at about 400 pps for the most basal electrodes, and about 200 pps for the most apical, which is qualitatively consistent with expectations based on the usually reported tonotopic response. These results should guide development of rate and place based coding strategies.

Jitter Sensitivity

The results showed that the thresholds of jitter discrimination were not dependent on the place of stimulation but were dependent on the rate of stimulation. Temporally jittered low rate pulse trains were easier to detect than jittered higher rate pulse trains. The thresholds ranged from on average 19% at 100pps to 34% at 500pps. This result constrains developments of rate and place based coding strategies.

Note Identification in Melodic Context

CIRs perform much worse than NH subjects in choosing the 'correct' note to complete a phrase from a well-known tune, erring by over a semitone on average and having a wide dispersion of choices.

Tonal Languages

Simulations of current CI speech coding strategies suggest that they are inadequate for the transmission of the tonal features of tonal languages. It was found that the identification rate of tonal words was near chance level when processed by regular speech processing strategies but improved if voice pitch information, F0, was included in the simulation. Speech processing strategies that can improve pitch discrimination would be extremely beneficial to CIRs who speak a tonal language who now rely almost solely on contextual cues for identification of these tonal words.

Music-L: Design and Results

The first suggestion based on the conclusions listed above is to devise new strategies. Accordingly, a strategy for combining rate and place information, based on the results above, was implemented. Preliminary results show improved pitch discrimination and perceived sound quality.

Pitch Discrimination

NH subjects score similarly when difference limens for change and frequency are measured using pure tone and complex stimuli. The results show that CIRs score better difference limens for change but worse difference limens for frequency using complex stimuli than when using a pure tone. Results of alternate pitch discrimination tests suggest using complex stimuli that CIRs perform well below that of NH subjects.

11.1 Suggestions for further research

11.1.1 Further Evaluation of Music-L

Although beyond the scope of this thesis, there is still much experimentation that can be done, as can be with any new processing strategy. A full evaluation of the Music-L strategies should be completed, and a training program that can compare these new strategies with regular speech processing strategies should be developed.

The rate pitch percept that the new strategy is intended to evoke should resemble that which occurs when the basilar membrane vibrates at low frequencies in response to a low frequency signal. For prelingually deafened CIRs the only stimulation they have ever received is via the implant. The pitch percept will then be based upon place, due to current speech coding strategies, and the rate-pitch percept will be a new concept. It is of interest to compare, not only whether both groups can improve pitch discrimination with Music-L, but also whether there is a difference in the amount of improvement between pre- and postlingually deafened recipients.

For NH children, the ability to discriminate pitch is dependent on age. It is of interest to compare the data gathered for NH children using a test such as the PMMA by Gordon, to those from children CIRs of similar age to compare how these groups differ not only in their absolute levels but also in their rate of improvement as a function of age. This experiment could also be performed with their normal speech processing strategy.

11.1.2 Proposals for future research in music

It appears that if music consisted of unaccompanied, monophonic sine waves, recipients might be able to follow the melody more clearly. Music, however, is performed by a range of instruments in solo and ensemble formats. To reduce the ambiguity in pitch direction that many CIRs experience may mean that improvements must be made to the peak picking strategies so that the harmonics picked do not confuse or misrepresent the pitch that attempt to convey. If, after selecting the maxima, post-processing was performed to ensure that the channels and pulses presented would represent the desired pitch, improvements may be obtained.

Another possible solution would be to construct a pitch estimate from channel and extracted frequencies information. This pitch could be represented by a pre-configured pattern of pulses that was known to elicit the correct pitch for each recipient.

To reinforce the perception of pitch higher frequency channels could also encode low frequency pitches by forced amplitude modulation.

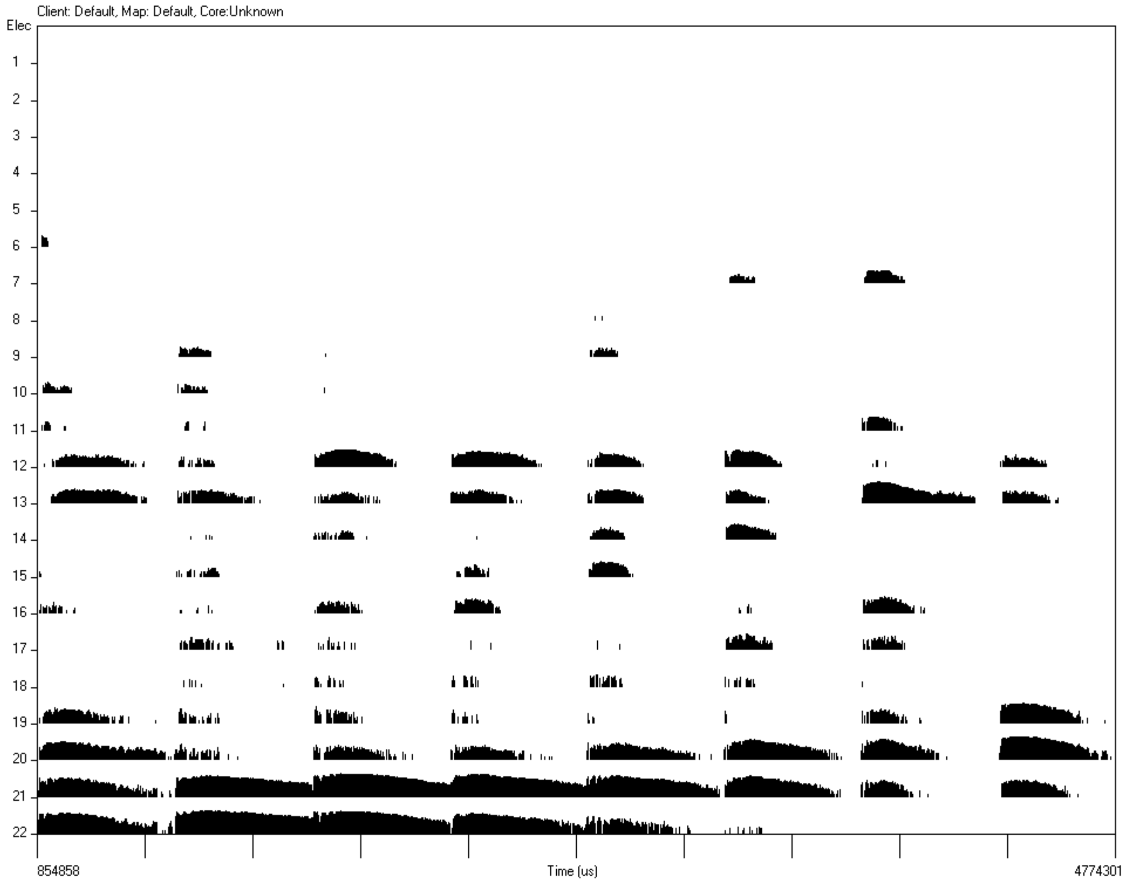
The experiments with temporal jitter revealed that for some subjects at some jitter threshold and rate/place combinations, a jittered pulse train was perceived to start smooth and then become rough at some point during the stimulation pattern even though there was temporal jitter throughout the pulse train. This may be due to an integration time required for the brain to realise that the temporal jitter is in fact present and not natural neural jitter. An interesting experiment would be to investigate the limits of the region in parameter space over within which this response is elicited. This may reveal how a sine wave can be perceived as pure even though the normal temporal code has naturally occurring neural jitter.

The results of this study show why existing strategies are poor at conveying musical information, and have determined some of the basic requirements of new strategies, which will improve perception of music and tonal languages in the future. The strategy developed here, and submitted to preliminary tests, is already yielding promising results and may be considerably improved when higher order complications and effects have been further investigated.

The results of this study show why existing strategies are poor at conveying musical information, and have determined some of the basic requirements of new strategies, which will improve perception of music and tonal languages in the future. The strategy developed here, and submitted to preliminary tests, is already yielding promising results and may be considerably improved when higher order complications and effects have been further investigated.

Appendix A

A C major scale was captured using the ACE program and displayed as an electrodogram, shown in Figure A-11-1. As the scale ascends from left to right one sees how each note would is represented through pulsatile stimulation. Of interest are the 3rd and 4th notes of the C major scale (E and F), which are a semitone apart. The resultant pulsatile stimulation is almost identical except for a small amount of extra stimulation by the note F on channel 15. The other semitone occurs between the 7th and 8th notes, from the B to C. There is more visible difference in the pulsatile stimulation between these two notes, because the frequency is higher so they fall in a range where the ACE map has more resolution.



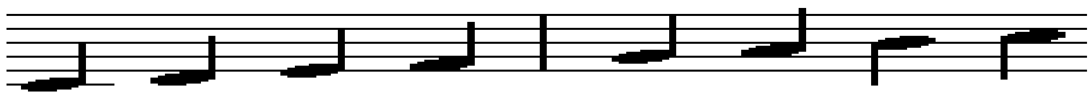


Figure A-11-1 Electrodogram of a C major scale beginning with middle C as recorded using an ACE filterbank map. The staff showing the C Major scale is also shown (x tick is ~0.4s).

An electrodogram was captured of the phrase used in the Happy Birthday test with the correct last note and is shown in Figure A-11-2. It can be seen that the last note does, under visual inspection, seem to have some ambiguity with regards to the direction of pitch. The 1st, 2nd and 4th notes, which are all the same note do appear higher in pitch if just examining the lowest 5 electrodes 22-18. These notes have strong activation of electrodes 20 and 19 whereas the 3rd and final note do not. When examining the harmonics activating electrodes 13 and above the 3rd note has higher stimulation especially on electrodes 9. The final note however, has strong stimulation on electrodes 12 and 13 but none higher.

This type of stimulation pattern is typical for presented music and is easy to see why pitch judgments may be impaired.

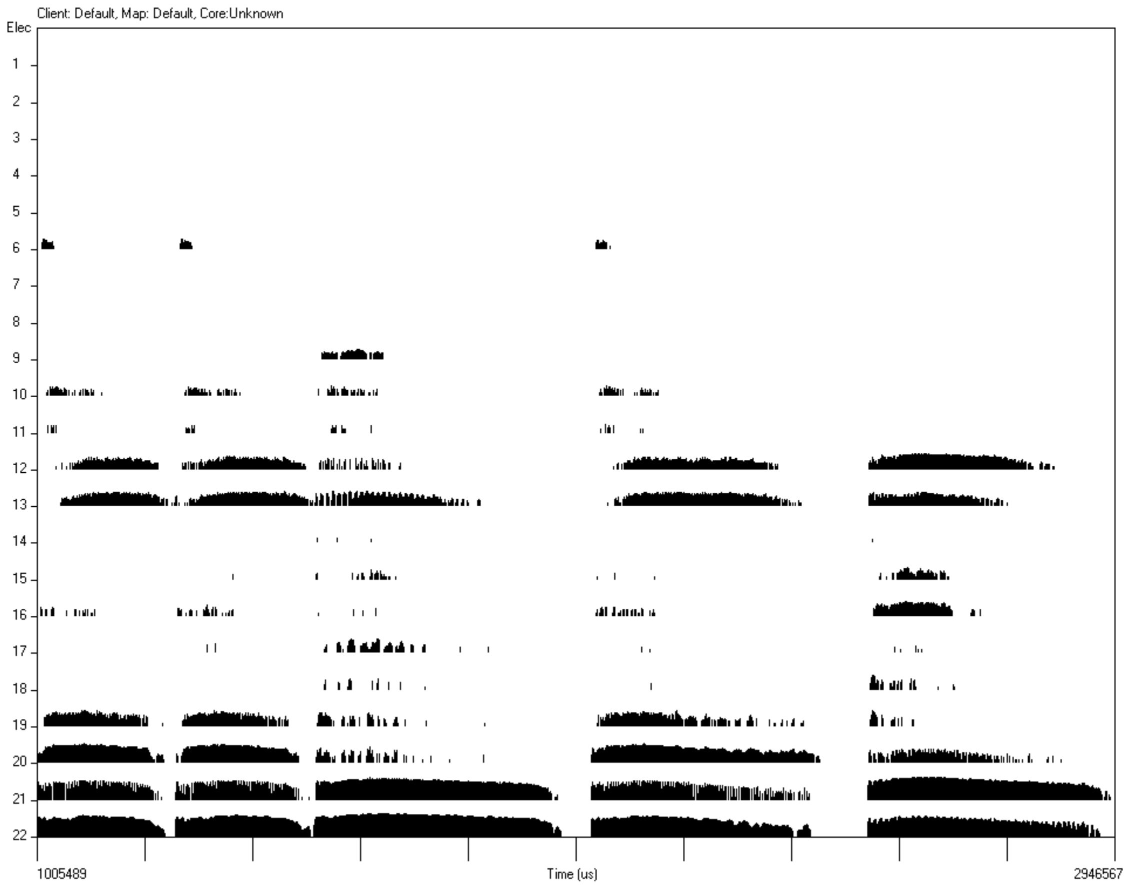


Figure A-11-2 Electrodogram of the phrase used in the Happy Birthday version of the test. This recording is with the correct last note (x tick is ~0.2s).

An electrodogram was captured for the phrase of Happy Birthday test to which the CIR were judging as correct. That is with the final note 3 semitones below the accepted note and equal to the 3rd note of presentation.

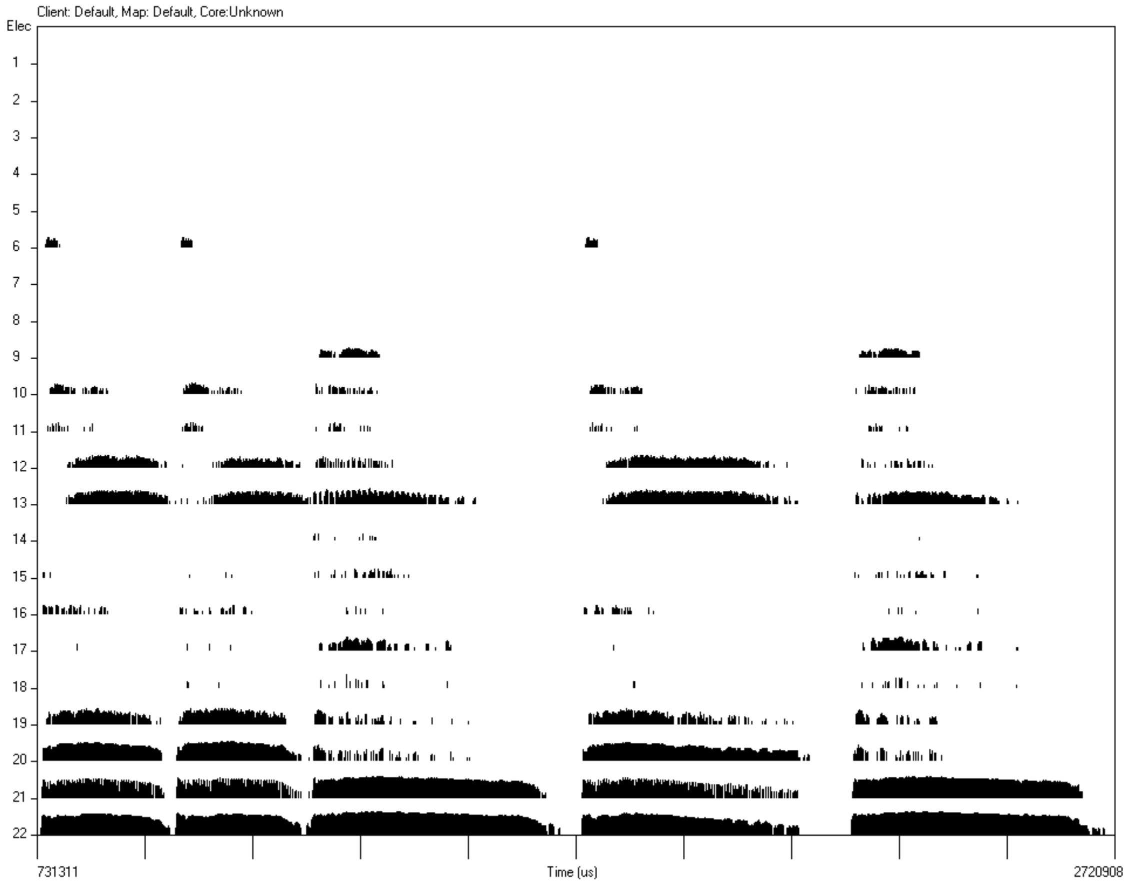


Figure A-11-3 Electrodogram of the phrase used in the Happy Birthday version of the test. This recording is with the last note that was selected by the CIR as the correct note (x tick is ~0.2s).

References

ANSI (1973). Psychoacoustical Terminology. New York, American National Standards Institute.

Antognelli, T., J. Patrick, P. Seligman, L. Whitford, C. M. Dobson and M. Follent (1991). "The 22 channel cochlear implant - an evolutionary tale." Australian Journal of Human Communication Disorders **19**(2): 6-12.

Aronson, L. and S. L. Arauz (1995). Fitting the Nucleus-22 Cochlear Implant for Spanish Speakers. International cochlear implant, speech and hearing symposium, Melbourne, Annals of Otology, Rhinology and Laryngology.

Bachem, A. (1948). "Chroma Fixation at the ends of the musical frequency scale." Journal of the Acoustical Society of America **20**(5): 704-705.

Bai, W. R. (1979). "Etymological notes on the Peking dialect." Fang Yan (Dialect): 171-177.

Barry, J., P. Blamey, K. Lee and D. Cheung (2000). Differentiation in tone production in Cantonese-speaking hearing impaired children. 6th International Conference on Spoken Language Processing,ICSLP, Inter-Speech 2000, Beijing, China.

Bekesy, G. v. (1960). Experiments in Hearing. New York, McGraw-Hill.

Bilsen, F. A. (1976). "Pitch of noise signals: Evidence for a "central spectrum"." Journal of the Acoustical Society of America **61**(1): 150-161.

Blamey, E. S. Parisi and G. M. Clark (1994). Pitch matching of electric and acoustic stimuli. International Cochlear Implants, Speech and Hearing Symposium, Melbourne, Annals of Otology, Rhinology and Laryngology.

Blamey, P., P. Arndt, F. Bergeron, G. Brdeberg, J. Brinacombe, G. Facer, J. Larky, B. Lindstrom, J. Nedzelski, A. Peterson, D. Shipp, S. Staller and L. Whitford (1996b). "Factors affecting auditory performance of postlinguistically deaf adults using cochlear implants." Audiology and Neuro-Otology **1**: 293-306.

Blamey, P. J., L. F. A. Martin and G. M. Clark (1985). "A comparison of three speech coding strategies using an acoustic model of a cochlear implant." Journal of the Acoustical Society of America **77**(1): 209-217.

Blanchfield, B. B., J. J. Feldman and j. L. Dunbar (1999). The severely to profoundly hearing impaired population in the United States: prevalence and demographics. www.cochlearimplants.com/research_general_brief.html**:** 4.

Braswell, C., A. Decuir, C. Hoskins, E. Kvet and G. Oubre (1988). "Relation between musical aptitude and intelligence among mentally retarded , advantaged , and disadvantaged subjects." Perceptual and Motor Skills **67**: 359-364.

Bredberg, G. and B. Lindstrom (1995). Insertion length of electrode array and its relation to speech communication performance and nonauditory side effects in multichannel-implanted patients. International Cochlear Implant, Speech and Hearing Symposium, Melbourne, Annals of Otology, Rhinology and Laryngology.

Brown, J. C. (1992). "Musical fundamental frequency tracking using a pattern recognition method." Journal of the Acoustical Society of America **92**(3): 1394-1402.

Bruce, I. (1997). Spatiotemporal coding of sound in the auditory nerve for cochlear implants. Otolaryngology. Melbourne, Melbourne**:** 189.

Busby, P. A. and G. M. Clark (1997). "Pitch and loudness estimation for single and multiple pulse per period electric pulse rates by cochlear implant patients." Journal of the Acoustical Society of America **101**(3): 1687-1695.

Busby, P. A., L. A. Whitford, P. J. Blamey, L. M. Richdarson and G. M. Clark (1994). "Pitch perception for different modes of stimulation using the Cochlear multiple-electrode prosthesis." Journal of the Acoustical society of America **95**(5 Pt 1 May): 2658-2669.

Cariani, P. (1996). Temporal coding of musical form. ICMPC, Montreal.

Chatterjee, M. (1999). "Effects of stimulation mode on threshold and loudness growth in multielectrode cochlear implants." Journal of the Acoustical Society of America **105**(2): 850-858.

Chatterjee, M. and J. J. Zwislocki (1997). "Cochlear mechanism of frequency and intensity coding." Hearing Research **111**: 65-67.

Chen, J. M., R. Farb, L. Hanusaik, D. Shipp and J. M. Nedzelski (1999). "Depth and quality of electrode insertion: A radiological and pitch scaling assessment of two cochlear implant systems." American Journal of Otology **20**(2): 192-197.

Ciocca, V., R. Aisha, A. Francis and L. Wong (2000). Can Cantonese children with Cochlear implants perceive lexical tones. 6th International Conference on Spoken Language Processing, ICSLP, Inter-Speech 2000, Beijeng, China.

Clark, G. M. (1992a). Cochlear Implants: future research directions. International Cochlear Implant, Speech and Hearing Symposium, Melbourne, Annals of Otology, Rhinology and Laryngology.

Clark, G. M. (1992b). "The development of speech processing strategies for the University of Melbourne/Cochlear multiple channel implantable hearing prosthesis." Journal of Speech-Language Pathology and Audiology **16**(2): 95-111.

Clark, G. M., P. J. Blamey, A. M. Brown, P. A. Busby, R. C. Dowell, N. K.-H. Franz, R. K. Shepherd, Y. C. Tong, R. L. Webb, M. S. Hirshorn, K. Kuzma, D. J. Mecklenburg, D. J. Money, J. F. Patrick and P. M. Seligman (1987). The University of Melbourne-Nucleus multiple electrode cochlear implant. Advances in Oto-Rhino-Laryngology. C. R. Pflatz. Basel, Karger.

Clark, G. M., Y. C. Tong and J. F. Patrick (1990). History of the Cochlear implant. Cochlear Prostheses. Melbourne, Churchill Livingstone**:** 1-5.

Clarkson, M. G. and E. C. Rogers (1995). "Infants require low-frequency energy to hear the pitch of the missing fundamental." Journal of the Acoustical Society of America` **98**(1): 148-154.

Cohen, M. A., S. Grossberg and L. L. Wyse (1995a). "A spectral network model of pitch perception." Journal of the Acoustical Society of America **98**(2(1)): 862-879.

Collins, L. M., T. A. Zwolan and G. H. Wakefield (1997). "Comparison of electrode discrimination, pitch ranking and pitch scaling data in postlingually deafened adult cochlear implant subjects." Journal of the Acoustical Society of America **101**(1 Jan): 440-455.

Cornsweet, T. N. (1962). "The staircase-method in psychophysics." American Journal of Psychology **75**: 485-491.

Cummins, F. (1998). Patterns in Human Language: pitch, tone, intonation II, http://gahu.ucd.ie/~fred/courses/phonetics/tone.html.

Darrow, A.-A. (1987). "An investigative study: The effect of hearing impairment on musical aptitude." Journal of Music Therapy **24**(2): 88-96.

Deutsch, D. (1992). "Paradoxes of musical pitch." Scientific American **August**: 70-75.

Dixon, W. J. and F. J. J. Massey (1983). The "Up-and Down" Method. Introduction to Statistical Analysis, McGraw Hill**:** 429-438.

Dixon, W. J. and A. M. Mood (1948). "A method for obtaining and analysing sensitivity data." Journal of the American Statistical Association **43**: 109-127.

Dobie, R. A. and N. Dillier (1985). "Some aspects of temporal coding for single-channel electrical stimulation of the cochlea." Hearing Research **18**: 41-55.

Dorman, Loizou and Rainey (1997). "Simulating the effect of cochlear-implant electrode insertion depth on speech understanding." Journal of the Acoustical Society of America **102**: 2993-2996.

Dorman, M., P. Loizou, J. Fitzke and Z. Tu (1999). "The recognition of monosyllabic words by cochlear implant patients and by normal-hearing subjects listening to words processed through cochlear implant signal processing strategies." Annals of Otology, Rhinology and Laryngology.

Dorman, M. F., M. Smith, L. Smith and J. L. Parkin (1994). "The pitch of electrically presented sinusoids." Journal of the Acoustical Society of America **95**(3 March): 1677-1679.

Doval, B. and X. Rodet (1991). Estimation of fundamental frequency of musical sound signals. Institute of Research and Coordination of Acoustics and Music, IRCAM, Paris, France, IEEE.

Dowell, R. C., Y. C. Tong, P. J. Blamey and G. M. Clark (1985). Psychophysics of multiple-channel stimulation. Cochlear Implants. R. A. Schindler and M. M. Merzenich. New York, Raven Press.

Durlach, N. I. and L. D. Braida (1969). "Intensity perception. I Preliminary theory of intensity resolution." Journal of the Acoustical Society of America **46**(2(2)): 372-383.

Eddington, D. K., W. H. Dobelle, D. E. Brackman, M. G. Mladejovsky and J. Parkin (1978b). "Place and periodicity pitch by stimulations of multiple scala tympani electrodes in deaf volunteers." Transactions of the American Society of Artificial Internal Organs **24**: 1-5.

Eddington, D. K., W. H. Dobelle, D. E. Brackman, M. G. Mladejovsky and J. L. Parkin (1978a). "Auditory Prostheses research with multiple channel intracochlear stimulation in man." Annals of Otology and Rhinology **87**(Supp 53): 5-39.

Evans, E. F. (1972). "The frequency response and other properties of single fibers in the guinea-pig cochlear nerve." Journal of Physiology **226**: 263-287.

Faulkner, A., S. Rosen and D. Stanton (2000c). Simulation of the effects of cochlear implant electrode insertion depth for tonotopically-mapped speech processors. London, Speech, Hearing and Language: work in progress No. 12, University of London.

Faulkner, A., S. Rosen and L. Wilkinson (2000a). Effects of the number of channels and speech to noise ration on rate of connected discourse tracking through a simulated cochlear implant speech-processor. London, Speech, Hearing and Language: work in progress No. 11, University College of London.

Fearn, R. (1998). The importance of F0 or voice pitch for perception of tonal language: Simulations with Cochlear implant speech processing strategies. 5th International Conference on Spoken Language Processing ICSLP '98, Sydney, Australia.

Fearn, R., P. Carter and J. Wolfe (1999). "The perception of pitch by users of cochlear implants: Possible significance for rate and place theories of pitch." Acoustics Australia **27**(2): 41-43.

Fearn, R. and J. Wolfe (2000). "Relative importance of rate and place: Experiments using pitch scaling techniques with cochlear implant recipients." Annals Of Otology, Rhinology and Laryngology **109 (12) pt 2**(Supp 185): 51-53.

Fingst, B. E., P. A. Burnett and D. Sutton (1983). "Intensity discrimination with cochlear implants." Journal of the Acoustical Society of America **73**(4): 1283-1292.

Fishman, K. E., R. V. Shannon and W. H. Slattery (1997). "Speech recognition as a function of the number of electrodes used in the SPEAK Cochlear implant speech processor." Journal of Speech, Language, and Hearing Research **40**: 1201-1215.

Fitch, J. and W. Shabana (1999). A wavelet-based pitch detector for musical signals. 2nd COST G-6 Workshop on Digital Audio Effects DAFx99, Trondheim, Norway.

Fletcher, H., Munson, W. (1933). "Loudness, its measurement and calculation." Journal of the Acoustical Society of America **5**: 82-108.

Fu, Q.-J., F.-G. Zeng, R. V. Shannon and S. D. Soli (1998). "Importance of tonal envelope cues in Chinese speech recognition." Journal of the Acoustical Society of America **104**(1): 505-510.

Gelfand, S. A. (1998). Hearing: An introduction to psychological and physiological acoustics. New York, Marcel Decker.

Gfeller, K., A. Christ, J. F. Knutson, S. Witt, K. T. Murray and R. S. Tyler (2000). "Musical backgrounds, listening habits, and aesthetic enjoyment of adult Cochlear implant recipients." Journal of the American Academy of Audiology **11**: 390-406.

Gfeller, K., J. F. Knutson, G. Woodworth and S. Witt (1998). "Timbral recognition and appraisal by adult cochlear implant users and normal hearing adults." Journal of the American Academy of Audiology **9**: 1-19.

Gfeller, K. and C. Lansing (1992). "Musical perception of cochlear implant users as measures by the Primary Measures of Music Audiation: An item analysis." Journal of Music Therapy **19**(1): 18-39.

Gfeller, K. and C. R. Lansing (1991). "Melodic, rhythm, and timbral perception of adult cochlear users." Journal of Speech and Hearing Research **34**: 916-920.

Gfeller, K., C. Turner, G. Woodworth, M. Mehr, R. Fearn, J. F. Knutson, S. Witt and J. Stordahl (2001). "Recognition of familiar melodies by adult cochlear implant recipients and normal hearing adults." Cochlear Implant International **In Press**.

Gfeller, K., S. Witt, K.-H. Kim and M. Adamek (1999). "A computerized music training program for adult cochlear implant recipients." Journal of the Academy of Rehabilitative Audiology **32**: 11-29.

Gfeller, K., G. Woodworth, D. A. Robin, S. Witt and J. F. Knutson (1997). "Perception of rhythmic and sequential pitch patterns by normally hearing adults and adult Cochlear Implant users." Ear and Hearing **18**(3): 252-260.

Gibbons, A. C. (1983a). "Primary Measures of Music Audiation scores in an institutionalised elderly population." Journal of Music Therapy **20**(1): 21-29.

Gibbons, A. C. (1983b). "Item analysis of the Primary Measures of Music Audiation in elderly care home residents." Journal of Music Therapy **20**(4): 201-210.

Goldstein, J. L. (1973). "An optimum processor theory for the central formation of the pitch of complex tones." Journal of the Acoustical Society of America **54**(6): 1496-1516.

Goldstein, M. J. H. (1994). "Auditory periphery as speech signal processor." IEEE Engineering in Medicine **April/May**: 186-196.

Gordon, E. E. (1979). "Developmental musical aptitude as measured by the Primary Measures of Music Audiation." Psychology of music **7**(1): 42-49.

Gordon, E. E. (1979). Primary Measures of Music Audiation- A music aptitude for kindergarten and primary grade children. Chicago, GIA Publications Inc.

Gordon, E. E. (1986). Intermediate Measures of Music Audiation, IMMA Manual.

Grant, P. (2000). Sensorineural Hearing Loss. www.medicineau.net.au/clinical/ent/SNHL.html**:** 15.

Green, D. M. (1970). Application and detection theory in psychophysics. Proceedings of the IEEE, IEEE.

Greenwood, D. (1990). "A cochlea frequency-position function for several species - 29 years later." Journal of the Acoustical Society of America **87**(6 June): 2592-2605.

Guttman, N. and S. Pruzansky (1962). "Lower Limits of Pitch and Musical Pitch." Journal of speech and hearing research **5**(5): 207-214.

Harris, J. D. (1952). "Pitch Discrimination." Journal of the Acoustical Society of America **24**(6 Nov): 750-755.

Hartmann, W. M. (1996). "Pitch, periodicity and the brain." Journal of the Acoustical Society of America **99**(4): 2457-2483.

Helmholtz, H. (1877). On the sensation of tone as a physiological basis for the theory of music- 2nd edition to conform to the 4th Edition of 1877, Dover Publications 1962.

Henry, K. R. (1996). "Auditory nerve neurophonic produced by the frequency difference of two simultaneously presented tones." Hearing Research **99**: 151-159.

Hiki, S., K. Imaizumi and Y. Fukuda (1998). Design of cochlear implant device for transmitting voice pitch information in speech sound of Asian languages. 5th International Conference on Spoken Language Processing ICSLP '98, Sydney, Australia.

Hoesel, R. J. M. v. and G. M. Clark (1997). "Psychophysical studies with two binaural cochlear implant subjects." Journal of the Acoustical Society of America **102**(1): 495-507.

Hose, B., G. Langner and H. Scheich (1987). "Topographic representation of periodicities in the forebrain of the mynah bird: one map for pitch and rhythm?" Brain Research **422**: 367-373.

House, W. F. and K. I. Berliner (1982). "Cochlear Implants: Progress in perspective." Annals of Otology, Rhinology, and Laryngology **91**: 124.

House, W. F. and J. Urban (1973). "Long term results of electrode implantation and electronic stimulation of the cochlea in man." Annals of Otolology **82**: 504-517.

Houtsma, A. J. M. and J. L. Goldstein (1971). "The central origin of the pitch of complex tones: Evidence from the musical interval recognition." Journal of the Acoustical Society of America **51**(2(2)): 520-529.

Huang, T.-S., N.-M. Wang and S.-Y. Liu (1995). Tone perception of mandarin-speaking postlingually deaf implantees using the Nucleus 22-channel cochlear mini system. International cochlear implant, speech and hearing symposium, Melbourne, Annals of Otology, Rhinology and Laryngology.

Huang, T.-S., N.-M. Wang and S.-Y. Liu (1996). "Nucleus 22 Channel cochlear mini-System implantations in Mandarin-speaking patients." The American Journal of Otology **17**: 46-52.

Irlicht, L., D. Au and G. M. Clark (1995). New temporal coding scheme for auditory nerve stimulation. International cochlear implant, speech and hearing symposium, Melbourne, Annals of Otology, Rhinology and Laryngology.

Jabri, M. A. and R. j. Wang, Eds. (1996). A novel channel selection system in cochlear implants using artificial neural networks. Advances in Neural Information Processing Systems, MIT Press.

Jesteadt, W. (1980). "An adaptive procedure for subjective judgments." Perception and Psychophysics **28**(1): 85-88.

Jones, P. (1994). The coding of voice-source information within a multichannel cochlear implant. Otolaryngology. Melbourne, Melbourne: 244.

Jules, F. M. G. (1995). Hearing aid including cochlear implant. France.

Klein, D. Zatorre, R,.J., Milner, Zhao, V., (2001) "Across linguistic PET study of Mandarin Chinese and English speakers." NeuroImage **13**: 646-657.

Kollmeier, B., R. H. Hilkey and U. K. Sieben (1988). "Adaptive staircase techniques in psychoacoustics: A comparison of human data and a mathematical model." Journal of the Acoustical Society of America **83**(5): 1852-1862.

Krumhansl, C. (1991). “Music psychology: Tonal structures in perception and memory.” Annual Review of Psychology **42**: 277-303.

Kral, A., R. Hartmann, D. Mortazavi and R. Klinke (1998). "Spatial resolution of cochlear implants: the electrical field and excitation of auditory afferents." Hearing Research **121**: 11-28.

Kwok, C. L., C. M. Wong, K. W. So, M. L. Yiu, C. C. Lau, W. S. Luk and S. O. Tang (1991). "Speech and lexical-tone perception in Cantonese-speaking implant patients." Australian Journal of Human Communication Disorders **19**(2): 77-90.

Lai, W. K., N. Dillier and H. Bogli (1993). "A Hybrid coding strategy for a multichannel cochlear implant." Adv Oyorhinolaryngology **48**: 91-96.

Langner, G. (1992). "Periodicity coding in the auditory system." Hearing Research **60**: 115-142.

Langner, G. (1997). "Neural processing and representation of periodicity pitch." Acta Otolaryngol (Stockh) **Supp 532**: 68-76.

Lawrence, M. and L.-G. Johnsson (1973). "The role of the organ of Corti in auditory nerve stimulation." Annals of Otolology **82**: 464-472.

Lawson, D., B. Wilson and M. zerbi (1999). Speech Processors for Auditory prostheses. NC, Quarterly Progress Report No 2, Research Triangle Institute.

Lawson, D. T., B. S. Wilson and M. Zerbi (1994). Speech Processors for auditory prostheses. NC, Quarterly Progress Report No 8, Research Triangle Institute.

Levitt, H. (1971). "Transformed up-down methods in psychoacoustics." Journal of the Acoustical Society of America **49**(2 pt 2): 467-476.

Licklider, J. C. R. (1951). "A duplex theory of pitch perception." Experientia **VII**: 128-134.

Licklider, J. C. R. (1954). "Periodicity pitch and place pitch." Journal of the Acoustical Society of America Journal of the Acoustical Society of America (26): 945-987.

Loeb, G. E., M. W. White and M. M. Merzenich (1983). "Spatial Cross correlation." Biol Cybern **47**: 149-163.

Loizou, P., M. Dorman and V. Powell (1998). "The recognition of vowels produced by men, women, boys and girls by Cochlear implant patients using a six-channel CIS processor." Journal of the Acoustical Society of America **103**(2): 1141-1149.

Loizou, P., M. Dorman, Z. Tu and J. Fitzke (1999). "The recognition of sentences in noise by normal-hearing listeners using simulations of SPEAK-type cochlear implant processors." Annals of Otology, Rhinology and Laryngology.

Loizou, P., S. Graham, J. Dickins, M. Dorman and O. Poroy (1997a). Comparing the performance of the SPEAK strategy (SPECTRA 22) and the CIS strategy (MED-EL) in quiet and in noise. 1997 Conference on Implantable Auditory Prostheses, Asilomar, Monterey CA.

Loizou, P. C. (1997b). Signal processing for cochlear prosthesis: A tutorial review. 40th MidWest symposium on Circuits and Systems, Sacramento, CA.

Loizou, P. C., M. F. Dorman, Z. Tu and J. Fitzke (2000). "Recognition of sentences in noise by normal-hearing listeners using simulation of speak-type cochlear implant signal processors." Annals of Otology, Rhinology and Laryngology **109 Number 12 pt 2**(Supp 185): 67-68.

Loizou, P. C. and O. Poroy (1999). A Parametric study of the CIS strategy. 1999 Conference on Implantable Auditory Prostheses, Asilomar, Monterey.

Maher, R. C. and J. W. Beauchamp (1994). "Fundamental frequency estimation of musical signals using a two-way mismatch procedure." Journal of the Acoustical Society of America **95**(4): 2254-2263.

McDermott, H. J. and C. M. McKay (1994a). "Pitch Ranking with non-simultaneous dual-electrode electrical stimulation of the cochlea." Journal of the Acoustical Society of America **96**(1 July): 155-162.

McDermott, H. J. and C. M. McKay (1994b). Musical pitch perception by users of the Nuclues 22 electrode cochlear implant. International cochlear implant and hearing symposium, Melbourne, Australia.

McDermott, H. J. and C. M. McKay (1997). "Musical pitch perception with electrical stimulation of the cochlea." Journal of the Acoustical Society of America **101**(3 March): 1622-1631.

McDermott, H. J., C. M. McKay and A. E. Vandali (1992a). "A new portable sound processor for the university of Melbourne/Nucleus Limited mulitelectrode cochlear implant." Journal of the Acoustical Society of America **91**(6): 3367-3371.

McKay, C. and H. McDermott (1997). Signal coding for cochlear implants: the effects of pulse width. 16th World congress of Otorhinolaryngology, head and neck surgery, Sydney, Monduzzi Editore.

McKay, C. M. and R. P. Carlyon (1999). "Dual temporal pitch percepts from acoustic and electric amplitude-modulated pulse trains." Journal of the Acoustical Society of America **105**(1): 347-357.

McKay, C. M. and H. J. McDermott (1996). "The perception of temporal patterns for electrical stimulation presented at one or two intracochlear sites." Journal of the Acoustical Society of America **100**(2 Pt 1 August): 1-11.

McKay, C. M., H. J. McDermott and G. M. Clark (1994a). "Pitch percepts associates with amplitude-modulated current pulse trains in cochlear implants." JASA **96**(5 Pt 1 Nov. 1994): 2664-2673.

McKay, C. M., H. J. McDermott and G. M. Clark (1994b). Loudness summation for the two channels of stimulation in cochlear implants: effects of spatial and temporal separation. International Cochlear Implants, Speech and Hearing Symposium, Melbourne, Annals of Otology, Rhinology and Laryngology.

McKay, C. M., H. J. McDermott and G. M. Clark (1995a). "Pitch matching of amplitude-modulated current pulse trains by cochlear implantees: The effect of modulation depth." Journal of the Acoustical Society of America **97**(3 March): 1777-1785.

McKay, C. M., H. J. McDermott and G. M. Clark (1996b). "The perceptual dimensions of single electrode and non-simultaneous dual-electrode stimuli in cochlear implantees." Journal of the Acoustical Society of America **99**: 1079-1090.

McKay, C. M., H. J. McDermott, A. Vandali and G. M. Clark (1992). "A comparison of speech perception of cochlear implantees using the spectral maxima sound processor (SMSP) and the MSP(MULTIPEAK) processor." Acta Otolaryngol (Stockh) **112**: 752-761.

Merzenich, M. M., R. P. Michelson, R. A. Schindler, C. R. Pettit and M. Reid (1973). "Neural encoding of sound evoked by electrical stimulation of the acoustic nerve." Annal Otol **82**: 486-503.

Moller, A. R. (1999). "Review of the roles of temporal and place coding of frequency in speech discrimination." Acta Otolaryngol (Stockh) **119**: 424-430.

Nobili, R. (1998). "How well do we understand the cochlear?" Trends in Neurosciences **21**(4): 159-167.

Ohmura, H. (1994). Fine pitch contour extraction by voicing fundamental. International Conference on Acoustics, Speech, and Signal Processing (ICASSP),, Adelaide, IEEE.

O'Leary, S. J., G. M. Clark and Y. C. Tong (1995). Discharge rate-level functions from dorsal cochlear nucleus single units in response to acoustic and electrical stimulation of the auditory nerve. International cochlear implant, speech and hearing symposium, Melbourne, Annals of Otology, Rhinology and Laryngology.

Online, N. C. (1995). Cochlear implants in adults and children. **1995:** 1-30.

Ouayoun, M., V. Pean, J. Genin, G. Bachelot, C. Fugain, B. Meyer and C.-H. Chourard (1997). "Asynchronous Interleaved Stimulation (AIS): A new speech coding strategy for cochlear implant." Acta Otolaryngol **117**: 182-186.

Peeters, S., A. Zarowski and F. E. Offeciers (1994). "New developments in hardware and software for cochlear implants." Polish journal of otolaryngology (Translated) **48**(Supp 15): 38-47.

Pfingst, B. E., L. A. Holloway, N. Poopat, A. R. Subramanya, M. F. Warren and T. A. Zwolan (1994). "Effects of stimulus level on nonspectral frequency discrimination by human subjects." Journal of hearing research **78**: 197-209.

Pfingst, B. E. and N. L. Rush (1987). Discrimination of simultaneous frequency and level changes in electrical stimuli. International cochlear implant sysmposium and workshop, Melbourne, Annals of Otology, Rhinology and Laryngolgoy.

Pijl, S. (1994). Musical pitch perception with pulsatile stimulation of single electrodes in patients implanted with the nucleus cochlear implant. International Cochlear Implants, Speech and Hearing Symposium, Melbourne.

Pijl, S. (1997a). "Pulse rate matching by cochlear implant patients: Effect of loudness randomization and electrode position." Ear and hearing research **18**(4): 316-325.

Pijl, S. (1997b). "Labeling of musical interval size by Cochlear Implant patients and normally hearing subjects." Ear and Hearing **October**.

Pijl, S. and D. W. F. Schwarz (1995a). "Intonation of musical intervals by deaf subjects stimulated with single bipolar cochlear implant electrodes." Hearing Research **89**: 203-211.

Pijl, S. and D. W. F. Schwarz (1995b). "Melody Recognition and musical interval perception by deaf subjects with electric pulse trains through single cochlear implant electrodes." Journal of the Acoustical Society of America **98**(2 Pt 1 August): 886-895.

Piszczalski, M. and B. A. Galler (1979). "Predicting musical pitch from component frequency ratios." Journal of the Acoustical Society of America **66**(3): 710-720.

Plomp, R. (1964). "The ear as a frequency analyzer." Journal of the Acoustical Society of America **36**(9).

Pollack, I. (1968a). "Detection and relative discrimination of auditory 'jitter'." Journal of the Acoustical Society of America **43**(2): 308-315.

Pollack, I. (1968b). "Discrimination of mean temporal interval within jittered auditory pulse trains." Journal of the Acoustical Society of America **43**(5): 1107-1112.

Pollack, I. (1968c). "Periodicity discrimination for auditory pulse trains." Journal of the Acoustical Society of America **43**(5): 1113-1119.

Pollack, I. (1971a). "Spectral Basis of auditory 'jitter' detection." Journal of the Acoustical Society of America **50**(2(2)): 555-558.

Pollack, I. (1971b). "Amplitude and time jitter thresholds for rectangular-wave trains." Journal of the Acoustical Society of America **50**(4(2)): 1133-1141.

Pollack, I. (1977). "Periodicity measures for repeated random auditory patterns." Journal of the Acoustical Society of America **63**(4): 1132-1144.

Pollack, I. (1979). "Discrimination of uniform spectrum pulse sequences." Journal of the Acoustical Society of America **66**(1): 115-122.

Rhode, W. S. (1998). "Neural encoding of single-formant stimuli in the ventral cochlear nucleus of the chinchilla." Hearing Research **117**: 39-56.

Pratt and P.E. Doak (1976). "A subjective rating scale for timbre." Journal of Sound and Vibration **45**: 317.

Repp, B.H. (1994). "The tritone paradox and the pitch range of the speaking voice: A dubious connection." Music Perception **12**: 227-255.

Rice, J. J., E. D. Young and G. A. Spirou (1995). "Auditory-nerve encoding of pinna-based spectral cues: Rate representation of high-frequency stimuli." Journal of the Acoustical Society of America **97**(3): 1764-1776.

Ritsma, R. J. and F. L. Engel (1964). "Pitch of Frequency-modulated signals." Journal of the Acoustical Society of America **36**(9).

Rosen, S., A. Faulkner and L. Wilkinson (1997). "Perceptual adaptation by normal listeners to upward shifts of spectral information in speech and its relevance for users of cochlear implants." Journal of the Acoustical Society of America **106**: 3629-3636.

Rosenberg, A. E. (1966). "Pitch discrimination of jittered pulse trains." Journal of the Acoustical Society of America **39**(5(1)): 920-928.

Schulz, E. and M. Kerber (1994). Music Perception with the MED-EL implants. Advances in Cochlear Implants. I. J. Hochmair-Desoyer and E. S. Hochmair, Manz Wein: 326-332.

Sek, A. and B. C. J. Moore (1995). "Frequency discrimination as a function of frequency, measured in several ways." Journal of the Acoustical society of America **97**(4): 2479-2486.

Seligman, P. (1985). Speech-processing strategies and their implementation. International cochlear implant symposium and workshop, Melbourne, Annals of Otology, Rhinology and Laryngology.

Shannon, R. (1983). "Multichannel electrical stimulation of the auditory nerve in man. I Basic Psychophysics." Hearing Research **11**: 157-189.

Shannon, R. (1992). "Temporal modulation transfer functions in patients with cochlear implants." Journal of the Acoustical Society of America **91**: 1974-1982.

Shannon, R. V., F. G. Zeng, V. Karmath, J. Wygonski and M. Ekelid (1995). "Speech recognition with primarily temporal cues." Science **270**(303-304).

Shannon, R. V., F. G. Zeng and J. Wygonski (1998). "Speech recognition with altered spectral distributions of envelope cues." Journal of the acoustical Society of America **104**: 2467-2476.

Shepard, R. N. (1964). "Circularity in judgments of relative pitch." Journal of the Acoustical Society of America **36**(12): 2346-2353.

Shiroma, M., T. Kikuchi, A. Kawano, M. Suzuki and K. Kaga (1997). Perception of music by cochlear implantee, http://www.rti.org/ciap97.pr-uu.htm.

Simmons, F. B., M. K. Herndon, L. E. Atlas, R. L. White and L. J. Dent (1984). Multielectrode modiolar stimulation: Some selected psychophysical and speech results. Advances in Audiology. W. D. Keidal and P. Finkenzeller. Basel, Karger. **2:** 163-169.

Stainsby, T. H., H. J. McDermott, C. M. McKay and G. M. Clark (1997). Preliminary results on spectral shape perception and discrimination of musical sounds by normal hearing subjects and cochlear implantees. Proceedings of the International Computer Music Conference (ICMC).

Stevens, S. (1935). "The relation of pitch to intensity." Journal of the Acoustical Society of America **6**(150-154).

Sutter, M. L. and C. E. Schreiner (1991). "Physiology and topography of neurons with multipeak tuning curves in cat auditory cortex." Journal of Neurophysiology **65**: 1207-1226.

Tang, S. O., W. S. Luk, C. C. Lau, K. W. So, C. M. Wong, M. L. Yiu and C. L. Kwok (1990). "Cochlear implant in Hong Kong Cantonese." The American Journal of Otology **11**(6): 421-426.

Taylor, M. M. and C. D. Creelman (1967). "PEST: Efficient Estimates on Probability Functions." Journal of the Acoustical Society of America **41**(4): 782-787.

Teich, M. C. and S. B. Lowen (1994). "Fractal patterns in auditory nerve-spike trains." IEEE Engineering in medicine and biology **April-May**: 197-201.

Terhardt, E. (1974). "Pitch, consonance, and harmony." Journal of the Acoustical Society of America **55**(5): 1061-1069.

Terhardt, E., G. Stoll and M. Seewann (1982). "Algorithm for extraction of pitch and pitch salience from complex tonal signals." Journal of the Acoustical Society of America **71**(3): 679-688.

Tong, Y. C., P. J. Blamey, R. C. Dowell and G. M. Clark (1983b). "Psychophysical studies evaluating the feasibility of a speech processing strategy for a multiple-channel cochlear implant." Journal of the Acoustical Society of America **74**(1): 73-80.

Tong, Y. C. and G. M. Clark (1985). "Absolute identification of electric pulse rates and electrode positions by cochlear implant patients." Journal of the Acoustical Society of America **77**(5 May): 1881-1888.

Tong, Y. C., R. C. Dowell, P. J. Blamey and G. M. Clark (1983a). "Two component hearing sensations produced by two-electrode stimulation in the cochlea of a deaf patient." Science **219**: 993-994.

Tong, Y. C., J. M. Harrison, J. Huigen and G. M. Clark (1990). "Comparison of two speech processing schemes using normal-hearing subjects." Acta Otolarnygol(stockh) **Supplement 469**: 135-139.

Townshend, B., N. Cotter, D. V. Compernolle and R. L. White (1987). "Pitch perception by cochlear implant subjects." Journal of the Acoustical Society of America **82**(1 July): 106-115.

Traux, B. (1999). Handbook for Acoustic Ecology, Cambridge Street Publishing.

Treutwein, B. (1995). "Adaptive Psychophysical Procedures." Vision Research **35**(17): 2503-2522.

Tye-Murray, N., M. Lowder and R. S. Tyler (1990). "Comparison of the F0F2 and F0F1F2 processing strategies for the cochlear corporation cochlear implant." Ear and Hearing **11**(3).

Tykocinski, M., R. K. Shepherd and G. M. Clark (1995). Acute effects of high-rate stimulation on auditory nerve function in guinea pigs. International cochlear implant, speech and hearing symposium, Melbourne, Annals of Otology, Rhinology and Laryngology.

Tyler, R. S., K. Gfeller and M. A. Mehr (2000). "A preliminary investigation comparing one and eight channels at fast and slow rates on music appraisal in adults with cochlear implants." Cochlear Implants International **1**(2): 82-87.

Vandali, A. E., L. A. Whitford, K. L. Plant and G. M. Clark (2000). "Speech Perception as a Function of Electrical Stimulation Rate: Using the Nucleus 24 Cochlear Implant System." Ear and Hearing **21**(6): 608-624.

Volta, A. (1800). "On the electricity excited by the mere contact of conducting substances of different kinds." Trans Roy Soc Phil **90**: 403-431.

von Bismarck, G. (1974). "Timber of steady sounds: A factorial investigation of its verbal attributes." Acustica **30**: 146-159.

Wallenberg, E. L. v. and R. D. Battmer (1991). "Comparative speech recognition results in eight subjects using two different coding strategies with the Nucleus 22 channel cochlear implant." British Journal of Audiology **25**: 371-380.

Wallenberg, E. L. v., R. D. Battmer, I. Doden, D. Gnadeberg, K. Hautle and T. Lenarz (1994). Place-pitch and speech perception measures with bipolar and Monopolar electrical stimulation of the cochlea. International Cochlear Implant, Speech and Hearing Symposium, Melbourne, Australia.

Wallenberg, E. L. v., E. S. Hocmair and I. J. Hochmair-Desoyer (1990). "Initial results with simultaneous analog and pulsatile stimulation of the cochlea." Acta Otolarnygol(stockh) **Supplement 469**: 140-149.

Wallenberg, E. L. v., R. Laszig, D. Gradeberg, R. Battmer, C. Desloovre, J. Kiefer, E. Lehnhardt and C. v. Ilberg (1994). Initial Findings with a modified nucleus implant comprised of 20 active intrcochlear and 2 extracochlear reference electrodes. Advances in Cochlear Implants. H. E. Hochmair-Desoyer IJ. Vienna, Austria, Manz**:** 186-192.

Walloch, R., R. Brummett, D. DeWeese and J. Vernon (1973). "Electrical stimulation of the inner ear." Annals of Otolology **82**: 473-485.

Wang, B. K. and T.-S. Huang (1988). "Current clinical results of the cochlear implant program conducted on mandarin-speaking patients." The American Journal of Otology **9**(1): 44-51.

Weinberger, N. M. and T. M. McKenna (1988). "Sensitivity of single neurons in auditory cortex to contour: toward a neurophysiology of music perception." Music Perception **5**: 355-390.

Wetherill, G. B. and H. Levitt (1965). "Sequential estimation of points on a psychometric function." The British Journal of Mathematical and Statistical Psychology **18**(1): 1-10.

Whitfield, I. C. (1980). "Auditory cortex and the pitch of complex tones." Journal of the Acoustical society of America **67**(2): 644-647.

Whitford, L. A., P. M. Seligman, P. J. Blamey, H. J. McDermott and J. F. Patrick (1993). "Comparison of current speech coding strategies." Adv Oyorhinolaryngology **48**: 85-90.

Whitford, L. A., P. M. Seligman, C. E. Everinhgam, T. Antognelli, M. C. Ckok, R. D. Hollow, K. Plant, E. S. Gerin, S. J. Staller, H. J. McDermott, W. Gibson and G. M. Clark (1995). "Evaluation of the nucleus spectra 22 processor and new speech processing strategt (SPEAK) in postlingually deafend adults." Acta Otolaryngol (Stockh) **115**: 1341.1-9.

Wightman, F. L. (1981). Pitch perception: an example of auditory pattern recognition. Auditory and Visual Pattern Recognition. J. D. J. Getty & J. H. Howard. Hillsdale, Erlbaum Press**:** 3-25.

Wilson, B., C. Finley, M. Zerbi, D. Lawson and C. v. d. Honet (1997b). Speech Processors for auditory prostheses. NC, RTI.

Wilson, B., D. Lawson, M. Zerbi and R. Wolford (1999). Speech processors for auditory prostheses. Iowa, 3rd Quarterly Progress Report, University of Iowa.

Wilson, B., M. Zerbi, C. Finley, D. Lawson and C. v. d. Honert (1997d). Speech Processors for Auditory Prostheses. NC, 8th Quartelry Progress Report, Research Triangle Institute.

Wilson, B. S., C. C. Finley, M. Zerbi and D. T. Lawson (1994). Speech processor for auditory prostheses. NC, 9th Quarterly Progress Report, Research Triangle Institute.

Wilson, B. S., D. T. Lawson, C. C. Finley and M. Zerbi (1991). Speech processor for auditory prostheses. NC, 10th Quarterly Progress Report, Research Triangle Institute.

Wong, S. L. (1984). A Chinese Syllabary Pronounced According to the Dialect of Canton. Hong Kong, Chung Hwa Book Co.

Xu, S. a., R. C. Dowell and G. M. Clark (1987). Results for Chinese and English in a multichannel cochlear implant patient. International cochlear implant symposium and workshop, Melbourne, Annals of Otology, Rhinology and Laryngology.

Yuen, A. F. C. Y. (1984). "The teaching of tones to children with profound hearing impairment." British Journal of Disorders of Communications **19**: 225-236.

Zeng, F.-G. (1999b). "Rate Discrimination and tone recognition in mandarin-speaking cochlear-implant listeners." Chinese Journal of Otorhinolaryngology **34**(2): 84-88.

Zeng, F.-G. and J. J. G. III (1999a). "Amplitude Mapping and Phoneme Recognition in Cochlear Implant Listeners." Ear and Hearing **20**(1): 60-74.

Zeng, F.-G. and R. V. Shannon (1992). "Loudness balance between electric and acoustic stimuli." Hearing Research **60**: 231-235.

Zeng, F.-G. and R. V. Shannon (1995). Loudness of simple and complex stimuli in electric hearing. International cochlear implant, speech and hearing symposium, Melbourne, Annals of Otology, Rhinology and Laryngology.

Zimmerman-Phillips, S. and C. Murad (1999). "Programming features of the Clarion multi-strategy cochlear implant." Annals of Otology, Rhinology and Laryngology **108 Supp 177**(4(2)): 17-21.

Zwislocki, J. J. (1991). "What is the cochlear place code for pitch." Acta Otolaryngol (Stockh) **111**: 256-262.

Zwislocki, J. J. (1995). Cochlear Precursors of neural pitch and loudness codes. International cochlear implant, speech and hearing symposium, Melbourne, Annals of Otology, Rhinology and Laryngology.

Zwolan, T. A., P. R. Kileny, C. Ashbaugh and S. A. Telian (1996). "Patient Performance with the Cochlear "20+2" Implant: Bipolar Versus Monopolar Activation." The American Journal of Otology **17**: 717-723.