Pitch Perception with Cochlear Implants

Brett Anthony Swanson

Submitted in total fulfilment of the requirements
of the degree of Doctor of Philosophy

August 2008

Department of Otolaryngology
The University of Melbourne
To Amanda
Abstract

Most cochlear implant recipients achieve good speech perception under good listening conditions. However, pitch perception is generally poor.

The first aspect of cochlear implant pitch perception investigated was place of stimulation. Sequential stimulation of neighbouring channels can produce pitch percepts intermediate to those of individual channels. This was evident when recipients pitch-ranked pure tones processed by the ACE strategy. The existing centroid model of place pitch perception was extended by incorporating a loudness model to provide a better fit to the experimental results.

The second aspect investigated was temporal cues on a single channel. A high-rate pulse train, modulated on and off at frequency F₀, had a higher pitch than a train of pulses at the rate of F₀. If amplitude modulation of high-rate pulse trains is to be used to convey pitch, then the shape of the modulating waveform is important: a half-wave shape is better than a square-wave (on-off) shape. The experimental results were consistent with a perceptual model that derived pitch from the times between successive auditory nerve firings (first-order inter-spike intervals). The results were not consistent with a model utilising auto-correlation or all-order inter-spike intervals.

Recipients showed limited ability to combine place and temporal cues to pitch.

In voiced speech, the harmonics are not resolved by the sound processor, and there are no useful place cues to pitch. Several experimental sound processing strategies that attempted to enhance temporal cues provided little benefit in laboratory testing. A strategy that used half-wave rectification (HWR) to provide temporal fine structure cues was implemented on the Freedom processor, but a take-home trial showed no significant difference in speech or pitch perception between HWR and the standard ACE strategy (which provided only envelope cues).

Pitch can be defined as that attribute of sensation whose variation is associated with melody. The Modified Melodies test was developed to measure pitch perception according to this definition. Normal hearing listeners had higher scores with harmonic tones than pure tones, and musicians performed better than non-musicians. Cochlear implant recipients performed much worse than subjects with normal hearing, but a contralateral hearing aid provided a large benefit.

Previous research had suggested that cochlear implant place pitch was more akin to brightness (an aspect of timbre) than to pitch. However, the Modified Melodies results supported the hypothesis that place pitch can support melody perception.
Declaration

This is to certify that

i. the thesis comprises only my original work towards the PhD except where indicated in the Preface below,

ii. due acknowledgement has been made in the text to all other material used,

iii. the thesis is less than 100,000 words in length

Preface

The study on contralateral hearing aids in chapter 11 was carried out by Penny Stewart for a Master of Audiology degree at the University of Melbourne, supervised by Cathy Sucher. I designed and implemented the Modified Melodies test, collaborated in the design of the protocol, and independently analysed the results.

The implementation of the HWR strategy on the Freedom processor in chapter 12 was based on the existing ACE implementation. I modified the assembly language code and performed verification, with assistance from Michael Goorevich, Tim Neal, Brionie Dayton, and Felicity Allen at Cochlear Ltd, Sydney. The protocol for the HWR take-home study was designed by Pam Dawson, and the recipients were tested by Pam Dawson and Michelle Knight at Cochlear Ltd, Melbourne. Pam Dawson analysed some of the results.
Acknowledgements

I could not have undertaken this thesis without the support of Cochlear Ltd, my employer since 1992, and in particular my co-supervisor Jim Patrick. My supervisor Hugh McDermott has been an inspiring mentor. My co-supervisor Colette McKay has been a source of wisdom on psychophysics. Peter Blamey guided this project through its early stages. Few PhD students have had supervisors with such a wealth of knowledge and experience.

I thank Cathy Sucher for the sung vowel stimuli (chapter 10); Penny Stewart for being the first external user of the Modified Melodies test and for sharing her results (chapter 11); Pam Dawson for running the HWR study in Melbourne (chapter 12) and for advice and feedback on the Modified Melodies test; Esti Nel for help with ethics applications and recruiting subjects; Ying-Yee Kong for providing her bimodal data (chapter 6); and Michael Goorevich, Tim Neal, Brionie Dayton, and Felicity Allen for help with Freedom programming (chapter 12). I am grateful to Bob Carlyon for reviewing material that became part of chapters 8, 9, and 10, and Doug Shaw for reviewing chapter 16. I have learned much from discussions with Andrew Vandali, Laurie Cohen, Richard Van Hoesel, Peter Busby, Peter Seligman, John Heasman, Bas van Dijk, Stoph Long, Zach Smith, Sean Lineaweaver, and Chris van den Honert.

I am grateful to all the people who volunteered to be subjects, particularly the implant recipients who surprised me with their generosity of time and effort.

My children, Jeremy and Bianca, helped me find and transcribe the melodies for the Modified Melodies test, and acted as enthusiastic subjects. Finally to my wife Amanda, I express my love and appreciation for all you have done, and I look forward to spending more time with you.
# Contents

Abstract .......................................................................................................................... iii  
Declaration ..................................................................................................................... iv  
Preface ........................................................................................................................... iv  
Acknowledgements ......................................................................................................... v  
Contents ......................................................................................................................... vi  
Figures ......................................................................................................................... viii  
Acronyms and Abbreviations......................................................................................... xii  
1 Introduction.................................................................................................................1  
   1.1 Background................................................................................. 1  
   1.2 Objectives ................................................................................. 1  
   1.3 Outline ........................................................................................ 1  
   1.4 Research questions ................................................................... 3  
2 Sound and hearing .....................................................................................................4  
   2.1 Introduction ............................................................................. 4  
   2.2 Pitch .......................................................................................... 4  
   2.3 Psychophysical procedures ....................................................... 4  
   2.4 Pitch perception ....................................................................... 12  
   2.5 Timbre perception ................................................................... 25  
   2.6 Vowel perception .................................................................... 28  
   2.7 Music perception ..................................................................... 29  
3 Cochlear implants .....................................................................................................33  
   3.1 Introduction .......................................................................... 33  
   3.2 Clinical aspects ....................................................................... 33  
   3.3 Cochlear implant systems ......................................................... 33  
   3.4 Electrical stimulation ............................................................. 34  
   3.5 Loudness .................................................................................. 37  
   3.6 Place pitch .............................................................................. 38  
   3.7 Temporal pitch ....................................................................... 41  
   3.8 Relationship between place and temporal cues ....................... 44  
   3.9 Conclusion .............................................................................. 48  
4 Cochlear implant sound processing .........................................................................49  
   4.1 Introduction .......................................................................... 49  
   4.2 Microphones and front end ..................................................... 49  
   4.3 Filterbank ............................................................................... 50  
   4.4 Envelope detection ................................................................. 53  
   4.5 FFT filterbank ........................................................................ 57  
   4.6 Responses to complex tones .................................................. 61  
   4.7 Sampling and selection ........................................................... 65  
   4.8 Amplitude mapping ................................................................. 66  
5 Cochlear implant pitch perception .............................................................................69  
   5.1 Introduction .......................................................................... 69  
   5.2 Testing pitch and melody perception ...................................... 69  
   5.3 Is place pitch really pitch? ...................................................... 77  
6 Improving cochlear implant pitch perception ..............................................................79  
   6.1 Introduction .......................................................................... 79
14.2 Cochlear implant place and temporal cues................................. 244
14.3 A new pitch perception model................................................. 245
14.4 Place pitch and melody............................................................ 247
14.5 Implications for cochlear implant sound processing................. 249

15 Appendix 1: Subjects.................................................................................. 252

16 Appendix 2: Statistics................................................................................... 254

16.1 Evaluating a single binomial experiment........................................ 254
16.2 Comparing two binomial experiments............................................ 256
16.3 Dependence of performance on a parameter................................. 265
16.4 Comparing two sets of binomial experiments......................... 270
16.5 Psychometric functions................................................................. 273
16.6 Comparing conditions having different parameter sets............ 274

17 References..................................................................................................... 277

Figures

Figure 2.1 A hypothetical psychometric function................................. 5
Figure 2.2 Probability Density Function of response to stimulus amplitude ................................. 6
Figure 2.3 Probability Density Functions for 2AFC signal detection task............................................ 6
Figure 2.4 Map reconstructed by MDS of 10 American cities............... 11
Figure 2.5 Cross-section of the cochlea.................................................... 12
Figure 2.6 Auditory model................................................................. 13
Figure 2.7 Lyon cochlear model (Slaney and Lyon 1990)......................... 14
Figure 2.8 Lyon filterbank frequency response..................................... 14
Figure 2.9 Lyon filterbank output for 200 Hz tones.............................. 16
Figure 2.10 Probability of auditory nerve firing for 200 Hz tones........... 17
Figure 2.11 Cochleograms for 200 Hz tones......................................... 18
Figure 2.12 Resolved and unresolved harmonics of a 200 Hz harmonic tone ............................ 18
Figure 2.13 Pitch perception model...................................................... 19
Figure 2.14 Average auditory nerve firing rate for 200 Hz tones............. 20
Figure 2.15 Frequency difference limens (Moore and Peters 1992)........ 21
Figure 2.16 Neuronal correlation mechanism (Licklider 1951)............. 23
Figure 2.17 Periodicity detector (Van Schaik 2001)............................... 24
Figure 2.18 Cochleograms for 200 Hz tones with lateral inhibitory network.......................... 25
Figure 2.19 Five of the harmonic tones used by Lichte (1941)............. 26
Figure 2.20 Pitch and brightness MDS study (Plomp 1976, chapter 6).... 28
Figure 2.21 Old MacDonald melody..................................................... 31
Figure 2.22 Old MacDonald contour.................................................... 31
Figure 2.23 Example of melodies used in Dowling and Fujitani (1971)..... 32
Figure 3.1 Nucleus Freedom cochlear implant system.......................... 34
Figure 3.2 Biphasic stimulation pulse waveform.................................... 36
Figure 3.3 Pulse sequence, showing current waveforms..................... 36
Figure 3.4 Pulse sequence, showing pulse timing and amplitudes........... 37
Figure 3.5 Three types of pulse sequence, all having a period of 10 ms... 42
Figure 3.6 Place and rate MDS experiment (Tong et al. 1983a)............ 45
Figure 3.7 Pitch scaling results (Eddington 1980, Figure 3)................ 46
Figure 3.8 Pitch scaling results (Fearn 2001, figure 3-4)....................... 47
Figure 3.9 Pitch scaling results (Zeng 2002, figure 2)......................... 48
Figure 9.1 Single-channel pulse sequences ..........................................................134
Figure 9.2 Single-channel modulation pitch ranking results: 6 semitone intervals .....141
Figure 9.3 Single-channel modulation pitch ranking results: 4 semitone intervals .....142
Figure 9.4 Single-channel modulation pitch ranking results: 1 semitone intervals .....142
Figure 9.5 Single-channel modulation pitch ranking group mean results ..........143
Figure 9.6 Pitch ranking result for subject S01 with MPPS sequences .................143
Figure 9.7 Psychometric functions for the pitch of 178 Hz MPPS sequence ..........147
Figure 9.8 Period of SPP sequence that matched pitch of 178 Hz MPPS sequence ..148
Figure 9.9 Warble discrimination stimuli: MPP sequences ....................................149
Figure 9.10 Warble discrimination stimuli: SPP sequences ....................................150
Figure 9.11 Warble discrimination results for 178 Hz SPP and MPP stimuli .........151
Figure 9.12 Warble detection thresholds (in μs) for SPP and MPP stimuli .............152
Figure 9.13 Summed autocorrelation of neural firing times for 178 Hz MPPS sequence. 155
Figure 10.1 Processing for one channel: ACE, HWG, TPS ....................................157
Figure 10.2 Pulse sequences for 157 Hz pure tone: ACE, HWG, TPS .................158
Figure 10.3 Pure tone pitch ranking results, ACE v HWG v TPS (6 semitone intervals)159
Figure 10.4 Pure tone pitch ranking results, ACE v HWG v TPS (4 semitone intervals)160
Figure 10.5 ACE electrodogram for four sung vowels (6 semitone intervals) .......165
Figure 10.6 Sung vowel pitch ranking results with ACE (6 semitone intervals) ....166
Figure 10.7 Sung vowel centroids (6 semitone intervals) .....................................167
Figure 10.8 Sung vowel centroid differences (6 semitone intervals) .....................167
Figure 10.9 Sung vowel pitch ranking results with ACE and Smoothed-ACE (6 semitone intervals) .................................................................168
Figure 10.10 Sung vowel pitch ranking results with ACE, HWR, TPS, and F0M (6 semitone intervals) .................................................................171
Figure 10.11 Sung vowel pitch ranking results with ACE, HWR, TPS, and F0M (1 semitone intervals) .................................................................172
Figure 11.1. The subject’s window of the Modified melodies software ................175
Figure 11.2. Old MacDonald ...............................................................................177
Figure 11.3. Twinkle Twinkle Little Star ..............................................................177
Figure 11.4. Frequency response of band-pass filter (green) and resulting line spectrum of 200 Hz harmonic tone (black) ....................................................178
Figure 11.5. Old MacDonald Stretch 1.11 distortion (red) overlayed on original (black)179
Figure 11.6. Twinkle Twinkle Stretch 1.11 distortion (red) overlayed on original (black)180
Figure 11.7. Old MacDonald Nudge 5 distortion (red) overlayed on original (black) ...181
Figure 11.8. Twinkle Twinkle Nudge 5 distortion (red) overlayed on original (black) ...181
Figure 11.9. Old MacDonald Exchange distortion (red) overlayed on original (black)183
Figure 11.10. Twinkle Twinkle Exchange distortion (red) overlayed on original (black)183
Figure 11.11. Old MacDonald (black - bottom) and Backwards distortion (red - top) ...184
Figure 11.12. Twinkle Twinkle (black - bottom) and Backwards distortion (red - top) ...185
Figure 11.13 Most familiar melodies (in alphabetical order) for Australian listeners ...189
Figure 11.14 Individual stretch results: 8-melody set, 262 Hz harmonic and pure tones191
Figure 11.15 Individual stretch results, 2-melody set, 262 Hz harmonic tones ........193
Figure 11.16 Group median stretch results, 2-melody set, 262 Hz harmonic tones ....194
Figure 11.17 Individual stretch results with 2-melody set, for 262 Hz harmonic tones and pure tones .................................................................195
Figure 11.18 Histogram of stretch 1.05 harmonic tone scores for musicians and non-musicians .................................197
Figure 11.19 Individual nudge results, 262 Hz harmonic tones ........................................199
Figure 11.20 Group median nudge results, 262 Hz harmonic tones ..................................200
Figure 11.21 CI individual stretch results, 262 Hz harmonic tones .....................................203
Figure 11.22 CI group median stretch results, 262 Hz harmonic tones ................................204
Figure 11.23 Individual stretch results: CI alone vs bimodal, 262 Hz harmonic tones ........208
Figure 11.24 Group median results: CI alone and Bimodal, 262 Hz harmonic tones ...........209
Figure 11.25 Scatter plot of summary scores of Bimodal vs CI alone .................................209
Figure 11.26 Contralateral ear audiograms ........................................................................212
Figure 11.27 Scatter plots of Modified Melodies summary scores against contralateral hearing thresholds ........................................................................................................213
Figure 11.28 Original (black) and stretch factor 4 (red) melodies ..................................214
Figure 11.29 Spectrograms of Old MacDonald .................................................................216
Figure 11.30 Spectrograms of Twinkle Twinkle Little Star ...............................................217
Figure 12.1 Freedom processor integrated circuit .........................................................220
Figure 12.2 Freedom processor software architecture ....................................................221
Figure 12.3 HWR processing .........................................................................................222
Figure 12.4 Electrodograms of ACE and HWR for a 262 Hz harmonic tone .....................223
Figure 12.5 ACE v HWR study protocol for subject S01 .....................................................225
Figure 12.6 Melbourne ACE v HWR study protocol ..........................................................226
Figure 12.7 Ascending melody (black), and Nudge 2 (red), Nudge 4 (blue) and Nudge 6 (pink) distortions .................................................................................................228
Figure 12.8 CNC word scores ACE v HWR .....................................................................229
Figure 12.9 Modified Melodies Stretch results, 149 Hz harmonic tones ............................230
Figure 12.10 Ascending Nudge results, 149 Hz harmonic tones ........................................232
Figure 12.11 Ascending Nudge group median results, 149 Hz harmonic tones ................233
Figure 12.12 Modified Melodies Nudge results, 149 Hz harmonic tones .............................235
Figure 12.13 Modified Melodies group median Nudge results, 149 Hz harmonic tones ....236
Figure 13.1 Old MacDonald electrodogram, pure tones, base frequency 523 Hz .............238
Figure 13.2 Old MacDonald, original and modified melodies ...........................................239
Figure 13.3 Individual results: place cues only (523 Hz pure tones) .................................241
Figure 13.4 Group median results: place cues only (523 Hz pure tones) ............................242
Figure 14.1 (a) Temporal pitch model using first-order inter-spike intervals (b) Modified model incorporating a lateral inhibitory network .........................................247
Figure 14.2 Using models in cochlear implants ...............................................................250
Figure 16.1 Probability density function for 2AFC with 8 trials .......................................256
Figure 16.2 Joint probability density function for the null hypothesis ..............................261
Figure 16.3 Joint probability density function showing the "extreme" outcomes ...............262
Figure 16.4 Joint probability density function showing the outcomes considered in Fisher's exact test ..........................................................264
Figure 16.5 Histogram of simulated scaled sample variance (black) and chi-square distribution (red) .........................................................................................................268
Figure 16.6 Modified Melodies summary scores for subject M02 .....................................275
## Acronyms and Abbreviations

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACE</td>
<td>Advanced Combinational Encoder (strategy)</td>
</tr>
<tr>
<td>ADC</td>
<td>Analog to Digital Converter</td>
</tr>
<tr>
<td>BP</td>
<td>Bipolar (stimulation mode)</td>
</tr>
<tr>
<td>C level</td>
<td>Maximum Comfortable level</td>
</tr>
<tr>
<td>CIS</td>
<td>Continuous Interleaved Sampling (strategy)</td>
</tr>
<tr>
<td>DSP</td>
<td>Digital Signal Processing</td>
</tr>
<tr>
<td>ECE</td>
<td>Extra-cochlear electrode</td>
</tr>
<tr>
<td>F0</td>
<td>Fundamental frequency</td>
</tr>
<tr>
<td>F0M</td>
<td>F0 Modulation (strategy)</td>
</tr>
<tr>
<td>FFT</td>
<td>Fast Fourier Transform</td>
</tr>
<tr>
<td>HWG</td>
<td>Half-wave Gating (strategy)</td>
</tr>
<tr>
<td>HWR</td>
<td>Half-wave rectification</td>
</tr>
<tr>
<td>LGF</td>
<td>Loudness Growth Function</td>
</tr>
<tr>
<td>MP</td>
<td>Monopolar (stimulation mode)</td>
</tr>
<tr>
<td>MPP</td>
<td>Multiple Pulse per Period (stimulation sequence)</td>
</tr>
<tr>
<td>MPPS</td>
<td>Multiple Pulse per Period – Synchronised (stimulation sequence)</td>
</tr>
<tr>
<td>MPPU</td>
<td>Multiple Pulse per Period – Uniform sampling (stimulation sequence)</td>
</tr>
<tr>
<td>NAFC</td>
<td>N alternative forced choice</td>
</tr>
<tr>
<td>NMT</td>
<td>Nucleus MATLAB Toolbox</td>
</tr>
<tr>
<td>PDT</td>
<td>Peak Derived Timing (strategy)</td>
</tr>
<tr>
<td>pps</td>
<td>Pulses per second</td>
</tr>
<tr>
<td>SPEAK</td>
<td>Spectral Peak (strategy)</td>
</tr>
<tr>
<td>SPP</td>
<td>Single Pulse per Period (stimulation sequence)</td>
</tr>
<tr>
<td>T level</td>
<td>Threshold level</td>
</tr>
<tr>
<td>TPS</td>
<td>Temporal Peak Sampling (strategy)</td>
</tr>
</tbody>
</table>
1 Introduction

1.1 Background

A cochlear implant restores a sense of hearing to a person with severe to profound deafness. More than 100,000 people have received a cochlear implant since they were introduced in the 1980s. With the latest devices, most cochlear implant recipients achieve good speech perception under good listening conditions. However, recipients sometimes describe the sound as unnatural, and pitch perception is generally poor. This limits their enjoyment of music, and causes difficulties with tonal languages.

1.2 Objectives

This thesis has the following objectives:

• To investigate cochlear implant place pitch and temporal pitch, evaluate existing perceptual models, and understand the relationship between these two aspects of perception.

• To develop a better method of testing cochlear implant pitch perception.

• To develop and evaluate a new sound coding strategy designed to improve cochlear implant pitch perception.

1.3 Outline

The thesis consists of two major parts.

Part 1 (chapters 2 – 6) reviews acoustic and electric hearing, with emphasis on pitch and music perception.

Chapter 2 is concerned with acoustic hearing. It examines the relationship between the physical parameters of sound and its perception (psychophysics), focussing on the perception of pitch, timbre, and melody.

Chapter 3 introduces cochlear implant systems, and how electrical stimulation is perceived.

Chapter 4 describes the sound processing implementation used in cochlear implant systems, concentrating on the Nucleus SPrint and Freedom processors.

Chapter 5 discusses test procedures for pitch and melody perception with cochlear implants.

Chapter 6 surveys the schemes that have been proposed to improve cochlear implant pitch perception.
Part 2 (Chapters 7 – 12) contains the experimental work of this thesis.

Chapter 7 documents the Nucleus MATLAB Toolbox research software developed by the author. It describes the strategies that were evaluated in this thesis, including Half-Wave Rectification (HWR), Half-Wave Gating (HWG), Temporal Peak Sampling (TPS), and F0 Modulation (F0M). It describes the psychophysical procedures used.

Chapter 8 explores cochlear implant place pitch perception. The ability of the ACE strategy to produce intermediate place pitch percepts was investigated. An existing model of place pitch perception was further developed to provide a better fit to the experimental results.

Chapter 9 studies cochlear implant temporal pitch perception, comparing the percepts generated by low-rate pulse trains and amplitude-modulated high-rate pulse trains. The required pulse timing accuracy of the implant system was investigated. Two models of temporal pitch perception were assessed for consistency with the experimental results.

Chapter 10 describes experiments where both place and temporal cues to pitch were available. Pitch-ranking ability was measured with the ACE, HWR, HWG, TPS, and F0M strategies. The results were compared to the previous chapters' results for place cues alone and temporal cues alone, to determine whether the subjects were able to usefully combine the two types of cue.

Chapter 11 follows the development of the new Modified Melodies test, and its validation with normal hearing listeners. It applies the Modified Melodies test to investigate cochlear implant pitch perception, both with and without a contralateral hearing aid.

Chapter 12 describes the implementation of the HWR strategy on the Nucleus Freedom processor. Speech perception and melody perception were evaluated after take-home use of the new strategy.

Chapter 13 addresses the question of whether cochlear implant place cues alone can support melody perception.

Finally, Chapter 14 provides a general discussion of the experimental results, summarises the findings of the thesis, and suggests avenues for further research.

Appendix 1 lists the biographical details of the cochlear implant subjects. Many of the subjects participated in multiple experiments, so their details are listed here instead of repeating them in each chapter. A cross-reference is provided showing which experiments they participated in.

Appendix 2 explains in detail the statistical methods used to analyse the experimental results.

It is assumed that the reader is familiar with basic signal processing concepts, such as spectral analysis and filters.
Cross-references to specific sections are denoted by the symbol "§".

1.4 Research questions

The experimental work in this thesis aimed to answer the following research questions concerning cochlear implant pitch perception:

- Are intermediate place pitch percepts generated by the ACE strategy?
- Are low-rate pulse trains and amplitude-modulated high-rate pulse trains equally effective in providing temporal pitch percepts?
- Does the pitch of an amplitude-modulated high-rate pulse train depend only on the fundamental period, or is the shape of the modulating waveform important?
- What stimulation timing resolution is required to provide good temporal pitch cues?
- Can recipients usefully combine place and temporal cues to pitch?
- Does a contralateral hearing aid improve cochlear implant pitch perception?
- Does the Half Wave Rectification (HWR) strategy improve pitch perception?
- Is place pitch really pitch – can it support melody perception?
2 Sound and hearing

2.1 Introduction

This chapter examines the perception of pitch, timbre, and melody in normal hearing. Some of the experimental methods used in the field are explained. Computational models of pitch perception are described.

2.2 Pitch

Pitch is a perceptual attribute of a sound, referring to the sensation it produces, rather than a physical attribute of the sound's waveform. Thus, measurements of pitch involve human subjects. This thesis will follow the definition of pitch used by Plack and Oxenham (2005a): "that attribute of sensation whose variation is associated with musical melodies".

The simplest sound that evokes a clear pitch is a pure tone, i.e. a sinusoid. Its pitch corresponds to its frequency. Pure tones are rarely heard in the environment. A complex tone is defined as a sum of sinusoids of different frequencies. As a special case, a harmonic tone is composed of sinusoids with frequencies that are integer multiples of the fundamental frequency, denoted \( f_0 \) (or F0), i.e.

\[
x(t) = \sum_n a_n \sin(2\pi nf_0 t + \phi_n)
\]  

(2.1)

The \( n \)th harmonic is the sinusoidal component indexed by \( n \) in Equation 2.1, with frequency \( nf_0 \), amplitude \( a_n \), and phase \( \phi_n \). The first harmonic, having frequency F0 (i.e. \( n = 1 \)), is called the fundamental. A harmonic tone is periodic, i.e. its waveform repeats with a period \( T_0 = 1/f_0 \). The pitch of a harmonic tone corresponds to its fundamental frequency.

2.3 Psychophysical procedures

Psychophysics is the study of how the perception of a stimulus is related to its physical parameters. Before describing pitch perception in more depth, it is useful to understand the methods that are used to measure perception. The experimental procedures used in this thesis will be described in more detail in §7.4.

2.3.1 Signal detection theory

A common psychophysical task in audiology is to measure a threshold, defined informally as the lowest amplitude signal that the subject can hear. A simple procedure is to present a pure
tone of a specific frequency at increasing levels, until the subject says that they can just hear it. If this task is repeated several times, it becomes clear that there is no single definitive value for the threshold. Instead, if the probability of the subject hearing the stimulus is plotted against sound pressure level (Figure 2.1), a curve is obtained, known as a psychometric function. The threshold can be defined as the level for which there is a 50% chance of a response. The threshold ideally should be a measurement of the subject's perceptual ability; however, the threshold obtained by this procedure is substantially affected by the instructions provided and the subject's attitude. Some subjects are reluctant to say that they heard the stimulus until they are certain, while others are more willing to respond even if they are unsure.

![Figure 2.1 A hypothetical psychometric function showing the dependence of the subject's response upon the stimulus level.](image)

The experimental procedure can be modified to include some trials where no stimulus is presented, but the subject is still asked whether they heard something. This is sometimes called a *yes/no* task, or a *single interval* task. The *hit rate* is the proportion of times the subject responded "yes" on those trials when the stimulus was presented. The *false alarm rate* is the proportion of times the subject responded "yes" on those trials when no stimulus was presented. Subjects will consciously or unconsciously assign differing importance to the two types of errors that can be made. For example, a subject may consider a false alarm to be a more serious mistake than missing a stimulus: in this case the subject is showing a preference for saying "no" when he is uncertain. This effect is known as subject bias. Two subjects with identical perceptual abilities can thus give very different hit rates.

Signal detection theory (Green and Swets 1989; McNicol 1972) provides a model of how subjects respond in these types of experiments. It assumes that each stimulus evokes a response on some internal perceptual scale. The response has two components: a deterministic component, dependent on the amplitude of the stimulus, and a random "noise" component. The response to a specific stimulus amplitude $a$ is thus modelled as a random variable $x$ with a probability density function centred at $x_a$ as shown in Figure 2.2. Signal detection theory
supposes that the subject establishes a criterion value \( x_c \), and responds whenever \( x > x_c \). A high criterion value lowers the false alarm rate, at the cost of a lower hit rate (because a larger signal is required for a positive response). The drawback of the yes/no task is that the researcher has limited influence on how the subject sets the criterion value.

Consider instead a two-alternative forced-choice (2AFC) task in which there are two observation intervals, of which only one (randomly selected) contains the signal. The subject's task is to select the interval that contained the signal. If the signal is too small for the subject to hear, he is forced to guess, and a mean score of 50% is expected. As the signal level is increased, the mean score should improve. The threshold can be defined as the level that produces a 75% score (midway between chance and ceiling). The benefit of this procedure is that it eliminates the subject's response bias: the two alternatives are both equally likely, and neither alternative is loaded with connotations of success or failure.

Each interval has an associated probability density function: the "no signal" interval is centred at \( x = 0 \), and the "signal" interval is centred at \( x = x_a \), as shown in Figure 2.3. The most effective decision strategy for a subject is to pick the interval that produced the largest response. Due to
the noise, occasionally the response to a "no signal" interval will be larger than the response to a "signal" interval, and so an error will be made. Although the actual units of the internal perceptual scale are not known, the probability of an error is clearly dependent on the relative sizes of the signal response and the noise. Conceptually, the internal perceptual scale can be normalised so that signal response amplitude is measured in units of the standard deviation of the noise. If it is assumed that the noise has a normal distribution, then the probability of correctly selecting the signal is (McNicol 1972):

\[
p = \Phi\left(\frac{d'}{\sqrt{2}}\right)
\]  

(2.2)

where \( \Phi \) is the cumulative normal distribution, and the sensitivity index \( d' \) is the difference in the mean responses to the two intervals on the normalised scale. The sensitivity index \( d' \) can be considered to be a measure of the perceptual difference between the two intervals, and is obtained by inverting Equation 2.2:

\[
d' = \sqrt{2} \Phi^{-1}(p)
\]  

(2.3)

Thus the sensitivity index \( d' \) can be estimated from the proportion correct in a 2AFC task as:

\[
d' = \sqrt{2} \Phi^{-1}\left(\frac{x}{n}\right)
\]  

(2.4)

where \( n \) is the number of trials, \( x \) is the number of correct responses, and \( \Phi^{-1} \) is the inverse of the cumulative normal distribution. This relationship applies to any 2AFC task, not just the signal detection task described here. If a subject obtains 100% correct on a block of trials, Equation 2.4 gives an infinite \( d' \) value. To avoid this, the formula can be modified (Trainor 1996):

\[
d' = \sqrt{2} \Phi^{-1}\left(\frac{x + \frac{1}{2}}{n + 1}\right)
\]  

(2.5)

It should be noted that an approximate estimate of the sensitivity \( d' \) can also be obtained from the hit rate and false alarm rate of a yes/no task (McNicol 1972), but the 2AFC relationship (Equation 2.4) is more accurate and requires fewer assumptions. Most experiments in this thesis used 2AFC tasks.

### 2.3.2 Discrimination and ranking

Another common psychophysical task is to investigate the perceptual difference between two stimuli. Terminology is not standardised in this area. In this thesis, a *ranking* task is defined as one that requires the subject to order the stimuli along a perceptual scale. For example, loudness
is perceived on a scale from soft to loud, so in a ranking task, the subject would hear two sounds and indicate which one was louder. In contrast, a discrimination task is defined as a task where the subject merely has to detect a difference between stimuli, without having to apply any ordering to them.

One type of discrimination task is the same/different task. In each trial, two stimuli are presented, and the subject is asked whether they are the same or different. In half of the trials, the two stimuli are the same (AA) and in the other half of the trials they are different (AB). The drawback is that a same/different task is equivalent (in signal detection theory) to a yes/no task, because the subject is asked to detect a difference (which may or may not be present). When the difference between the stimuli is small and the subject is uncertain, he may show a tendency to report "no difference". In other words, a same/different task suffers from subject response bias.

Just as in the threshold task (§2.3.1), a preferred form of discrimination task is the N-alternative forced choice (NAFC) task. In each trial, N stimuli are presented sequentially, of which one is different (e.g. ABAA). N is usually 3 or 4. The subject's task is to select the variant stimulus, which is randomly positioned on each trial. The chance score is 1/N.

Compared to ranking tasks, discrimination tasks suffer from two important limitations. Firstly, it is hard to rule out the possibility that the subject is using unintended cues to distinguish between the stimuli. Secondly, a discrimination task provides no evidence that the perceptual change is in the same direction as that expected by the researcher. Although rarely an issue when testing normal hearing subjects, this can be a concern with cochlear implant recipients, and will be further discussed in §5.2. Discrimination tasks are best suited to cases where the researcher is not certain how the subject will perceive the difference between the stimuli, or where there are no well-established labels for the percept. An example is the warble discrimination experiment that will be described in §9.4.

### 2.3.3 Methods for adjusting stimulus parameters

A psychophysics procedure often consists of a sequence of trials of a specific task, where some physical parameter of the stimulus is varied from trial to trial. To illustrate, consider using a 2AFC task to measure a threshold (§2.3.1). The objective is to find the level that gives a 75% probability of a correct response. In the method of constant stimuli (Levitt 1971), a fixed set of levels are specified, spanning the likely range of subject thresholds. Each trial is performed with a level chosen pseudo-randomly from the set, until a predetermined number of trials have been done at each level. If the proportion correct is plotted as a function of level, a psychometric function is obtained. The level that would give a score of 75% correct can be found by interpolation, or by a maximum-likelihood psychometric fitting algorithm as used in this thesis (§16.5) (Wichmann and Hill 2001). The method of constant stimuli has two disadvantages:
firstly, it requires some prior knowledge of the range of threshold levels. Secondly, it is inefficient, as trials conducted at levels far above or far below the threshold provide little information.

The alternative is an *adaptive* procedure, where the responses to previous trials determine the level used for the next trial. In the *up-down* method (Levitt 1971), the stimulus level is decreased after a correct response, and increased after an incorrect response, so that the level converges to the threshold. Because trials are concentrated in the region near the threshold, the efficiency is improved (i.e. fewer trials are needed). One drawback of this simple rule is that its nature often becomes apparent to the subject, which can influence the results. To avoid this, two independent instances of the rule (with different starting points) can be run concurrently, and trials can be randomly alternated between the two runs (Levitt 1971). To converge to a point on the psychometric function other than the midpoint, the adaptive rule can be modified: for example two consecutive correct responses may be required before the level is decreased (Levitt 1971). These are sometimes referred to as *n-up m-down* rules. Taylor and Creelman (1967) developed an adaptive procedure called PEST, in which a block of trials is conducted at each level.

The speed of convergence of these adaptive procedures is determined by the step size. Usually the step size is initially large, and is decreased after several changes in direction ("reversals"). A simple algorithm for calculating the final threshold estimate is to average the levels over the last set of reversals. Alternatively, the proportion correct at each level can be calculated, and a psychometric function can be fit using a maximum-likelihood algorithm (Hall 1981). This is more accurate as it uses information from all trials.

Maximum-likelihood techniques can be further utilised by making a new estimate of threshold after every trial, and using this estimate as the level for the next trial (Hall 1968; Watson and Pelli 1983; Harvey 1986). A further refinement is the QUEST method (Watson and Pelli 1983; King-Smith et al. 1994), which can also employ any available prior knowledge of the threshold level (e.g. obtained from previous studies with other subjects). The QUEST adaptive rule was used for loudness balancing in this thesis (§7.4.1).

### 2.3.4 Magnitude estimation

Stevens (1986) was the pioneer of the magnitude estimation approach, sometimes called magnitude scaling. A set of stimuli is defined that vary in physical magnitude. In each trial, a stimulus is randomly selected from the set and presented, and the subject is asked to provide a numerical estimate of its perceptual magnitude. For example, the subject hears a 3000 Hz pure tone, with amplitude chosen randomly on reach trial, and gives a numerical estimate of its loudness. Stevens conducted magnitude scaling experiments on a wide variety of human senses,
and postulated that in every case the perceived magnitude $\Psi$ of a sensation was related to the physical magnitude $I$ of the stimulus by a power law (Stevens 1986):

$$\Psi = k I^\beta$$

(2.6)

For example, the perceived loudness of a 3000 Hz pure tone was related to its sound pressure level by a power law with exponent of approximately $\beta = 0.6$. Stevens made a distinction between two types of sensory attribute (Stevens 1986, p 13):

The prototypes of the two kinds of perceptual continua are exemplified by loudness and pitch. Loudness is an aspect of sound that has what can best be described as degrees of magnitude or quantity. Pitch does not. Pitch varies from high to low; it has a kind of position, and in a sense it is a qualitative continuum. Loudness may be called a *prothetic* continuum, and pitch a *metathetic* one.

The power law only applied to prothetic attributes (such as loudness), and not metathetic attributes (such as pitch).

The magnitude estimation procedure has been criticised as susceptible to bias, with the results affected by extraneous factors such as whether a reference stimulus is provided, the instructions given, the range of stimuli presented, and the numerical range of estimates permitted or expected (Moore 1997). The results from individual subjects often differ widely, and it is necessary to average over a large number of subjects to obtain consistent results. The magnitude estimation procedure was not used in the experimental work of this thesis.

### 2.3.5 Multidimensional scaling

Multidimensional scaling (MDS) is a procedure used to investigate and visualise the perceptual similarities between a set of stimuli. To illustrate the principle, Table 2.1 lists the distances, in miles, between ten American cities (taken from the MATLAB Statistics Toolbox documentation for MDS). Given only this matrix of distances, the MDS algorithm can reconstruct a map of the city locations (Figure 2.4). The orientation of the reconstructed map is arbitrary, as the distance matrix contains no latitude or longitude information; in this case rotating the map by about 10 degrees clockwise gives the correct orientation. The *multidimensional* term refers to the algorithm's ability to fit the data to a geometric space with one, two, three or more dimensions. The *stress* is a measure of how well the distance matrix fits a specified number of dimensions. In this example, the stress is very high for a one-dimensional space (indicating a poor fit), and substantially lower for a two-dimensional space. Interestingly, a three-dimensional space gives a better fit, as the curvature of the earth is evident in the distance matrix.
Table 2.1 MDS data set: distances (in miles) between 10 American cities

<table>
<thead>
<tr>
<th></th>
<th>Atlanta</th>
<th>Chicago</th>
<th>Denver</th>
<th>Houston</th>
<th>LA</th>
<th>Miami</th>
<th>NY</th>
<th>SF</th>
<th>Seattle</th>
<th>Wash</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atlanta</td>
<td>0</td>
<td>587</td>
<td>1212</td>
<td>701</td>
<td>1936</td>
<td>604</td>
<td>748</td>
<td>2139</td>
<td>2182</td>
<td>543</td>
</tr>
<tr>
<td>Chicago</td>
<td>587</td>
<td>0</td>
<td>920</td>
<td>940</td>
<td>1745</td>
<td>1188</td>
<td>713</td>
<td>1858</td>
<td>1737</td>
<td>597</td>
</tr>
<tr>
<td>Denver</td>
<td>1212</td>
<td>920</td>
<td>0</td>
<td>879</td>
<td>831</td>
<td>1726</td>
<td>498</td>
<td>949</td>
<td>1021</td>
<td>1494</td>
</tr>
<tr>
<td>Houston</td>
<td>701</td>
<td>940</td>
<td>879</td>
<td>0</td>
<td>1374</td>
<td>968</td>
<td>1420</td>
<td>1645</td>
<td>1891</td>
<td>1220</td>
</tr>
<tr>
<td>LA</td>
<td>1936</td>
<td>1745</td>
<td>831</td>
<td>1374</td>
<td>0</td>
<td>2339</td>
<td>2451</td>
<td>347</td>
<td>959</td>
<td>2300</td>
</tr>
<tr>
<td>Miami</td>
<td>604</td>
<td>1188</td>
<td>1726</td>
<td>968</td>
<td>2339</td>
<td>0</td>
<td>1092</td>
<td>2594</td>
<td>2734</td>
<td>923</td>
</tr>
<tr>
<td>NY</td>
<td>748</td>
<td>713</td>
<td>1631</td>
<td>1420</td>
<td>2451</td>
<td>1092</td>
<td>0</td>
<td>2571</td>
<td>2408</td>
<td>205</td>
</tr>
<tr>
<td>SF</td>
<td>2139</td>
<td>1858</td>
<td>949</td>
<td>1645</td>
<td>347</td>
<td>2594</td>
<td>2571</td>
<td>0</td>
<td>678</td>
<td>2442</td>
</tr>
<tr>
<td>Seattle</td>
<td>2182</td>
<td>1737</td>
<td>1021</td>
<td>1891</td>
<td>959</td>
<td>2734</td>
<td>2408</td>
<td>678</td>
<td>0</td>
<td>2329</td>
</tr>
<tr>
<td>Wash</td>
<td>543</td>
<td>597</td>
<td>1494</td>
<td>1220</td>
<td>2300</td>
<td>923</td>
<td>205</td>
<td>2442</td>
<td>2329</td>
<td>0</td>
</tr>
</tbody>
</table>

Figure 2.4 Map reconstructed by MDS of 10 American cities

In a psychophysics application, each element of the matrix is a measure of the dissimilarity of a pair of stimuli (i.e. the perceptual distance between the stimuli). A straightforward procedure (e.g. McKay et al. 1996) for obtaining each dissimilarity value is to present the relevant pair of stimuli and ask the subject to rate them on a scale from indistinguishable (dissimilarity value of zero) to extremely different (dissimilarity value of 100). Alternatively, in the method of triadic comparison (e.g. Plomp 1976, p 87; Tong et al. 1983a; Tong et al. 1983b), three stimuli from the set are presented to the subject in each trial. The dissimilarity matrix element corresponding to the pair of stimuli which are judged by the subject to be most different is incremented by two, the element for the pair which are judged as most similar is left unchanged, and the element for the remaining pair is incremented by one.

The advantage of multidimensional scaling is that the relationships that are found between the stimuli are derived entirely from the subjects' judgements, independent of any preconceived
2.4 Pitch perception

2.4.1 The ear

The ear consists of three parts: the outer ear, which aids sound localisation; the middle ear, which efficiently transfers sound energy to the inner ear; and the inner ear, which contains the cochlea and the vestibular (balance) system. The cochlea is a spiral-shaped, fluid-filled chamber of bone. A cross-section is shown in Figure 2.5. The scala media and the scala tympani are separated by the basilar membrane, which supports the organ of Corti. The basilar membrane vibrates in response to sound. The mechanical properties of the basilar membrane vary progressively along its length, such that each position has a characteristic frequency to which it has the highest sensitivity. The basal end is most sensitive to high frequencies, and the apical end to low frequencies.

Figure 2.5 Cross-section of the cochlea

Basilar membrane motion is sensed by the inner hair cells. Deflections of the basilar membrane towards the tectorial membrane cause the inner hair cells to excite neurons known as spiral ganglion cells. No excitation occurs for deflections in the opposite direction, thus the nerve firings carry information on the timing of the basilar membrane motion at each place. This phase-locking mechanism is effective for audio frequencies up to about 5 kHz. The nerve fibres (axons) of the spiral ganglion cells are bundled together in the auditory nerve, and carry information to the brain.
2.4.2 Auditory models

Many signal processing models of the ear have been proposed, embodying a range of trade-offs between simplicity and faithful reproduction of physiological responses. The models generally have the structure shown in Figure 2.6.

![Figure 2.6 Auditory model](image)

The outer and middle ears are often modelled by a simple band-pass filter, most sensitive to the frequency range 500 – 4000 Hz. Each place on the basilar membrane is modelled as a band-pass filter centred at its characteristic frequency. The bandwidth of each such auditory filter is roughly proportional to its characteristic frequency.

The transduction stage models the inner hair cells and the generation of action potentials in the spiral ganglion cells. The output of the model is a representation of auditory neuron firing as a function of place and time.

One computationally efficient approach is the Lyon cochlear model (Lyon 1982; Lyon 1983; Slaney and Lyon 1990; Slaney and Lyon 1993). A MATLAB implementation of the model, written by Malcolm Slaney (Slaney 1998), was used in this thesis (§7.2). The filterbank is implemented as a cascade of second-order filters, as shown in Figure 2.7. The frequency response at each tap along the cascade is asymmetric, with a very sharp high-frequency cut-off, as shown in Figure 2.8. In this example, there were 173 sections, but for clarity, only every 8th tap is shown. The spacing of the filter centre frequencies is approximately linear below 1 kHz, and logarithmic above 1 kHz (note the logarithmic frequency axis in Figure 2.8).

The inner hair cells are modelled by half-wave rectifiers. The filterbank itself is linear; the non-linearity of the basilar membrane response is modelled by a coupled AGC stage following the rectifiers. The output signal represents the time-varying probability of firing of each auditory neuron.
Despite its simplicity, the Lyon cochlear model possesses many attributes that are important in pitch perception. Three tones, each having a fundamental frequency of 200 Hz, will be used to illustrate: a pure tone; a tone comprising harmonics 1:32; and a tone comprising harmonics 12:32. The filterbank response to each tone is shown in Figure 2.9, for the six taps that have characteristic frequencies of approximately 200, 400, 800, 1600, 3200, and 6400 Hz (i.e. the filters that are centred on harmonics 1, 2, 4, 8, 16 and 32). The model output (the spike probability) for the same six channels is shown in Figure 2.10. In this example, the AGC of Figure 2.7 was omitted, and the half-wave rectifier output was smoothed by a 1000 Hz low-pass filter to model the loss of phase locking at high frequencies. Cochleograms for the three tones are shown in Figure 2.11. A cochleogram is an alternative method of displaying the model output as a function of time and place. The outputs of all channels are displayed as an image, with brightness indicating amplitude. The vertical axis represents cochlear place, and is labelled with the characteristic frequency of the corresponding filter.
The cochleogram for the 200 Hz pure tone (Figure 2.11a) demonstrates the travelling wave concept. Proceeding from the basal end, the response gradually grows in amplitude until it reaches a peak at the 200 Hz place. More apical places show no response. Each filter that does respond produces a 200 Hz sinusoid, but with a gradually increasing phase shift. The largest phase shift occurs where the amplitude peaks. This rapid phase transition has been observed in animal studies (Shamma 1985a).

Because the auditory filter bandwidths increase with frequency, while harmonics are spaced linearly with frequency, the response to a harmonic tone is more complicated. Figure 2.12 shows the first 32 harmonics of 200 Hz on a linear frequency scale. It also shows the responses of two auditory filters. The 400 Hz filter, centred on the second harmonic, is sufficiently narrow that it has little response to any other harmonic. In contrast, the 3200 Hz filter, centred on the 16th harmonic, is much wider and responds to many harmonics. Whenever more than one component falls within the pass-band of a filter, the amplitude modulates at the fundamental frequency. This can be seen in the filterbank response to the tone containing the first 32 harmonics (Figure 2.9b). The filter at the 200 Hz place responds only to the fundamental. The 400 Hz place responds mainly to the second harmonic; its output is approximately a 400 Hz sinusoid. The response becomes less sinusoidal and more modulated at the basal end. The lower harmonics are said to be resolved by the filterbank; conversely the upper harmonics are unresolved. Examining Figure 2.10b and Figure 2.11b, it can be seen that at the apical end, the nerve firing is phase-locked to the individual resolved harmonics. At the basal end, the nerve firing follows the envelope, and so modulates at the fundamental frequency.

Finally, consider the tone consisting of only the upper harmonics. The responses are restricted to the basal half of the cochlear, where the harmonics are unresolved. The neural response again modulates at the fundamental frequency (Figure 2.10c and Figure 2.11c).
Figure 2.9 Lyon filterbank output for 200 Hz tones.
(a) pure tone; (b) tone with harmonics 1:32; (c) tone with harmonics 12:32
Figure 2.10 Probability of auditory nerve firing for 200 Hz tones. 
(a) pure tone; (b) tone with harmonics $1:32$; (c) tone with harmonics $12:32$
Figure 2.11 Cochleograms for 200 Hz tones.
(a) pure tone; (b) tone with harmonics 1:32; (c) tone with harmonics 12:32

Figure 2.12 Resolved and unresolved harmonics of a 200 Hz harmonic tone.
The frequency responses of the auditory filters centred at harmonic 2 (400 Hz) and harmonic 16 (3200 Hz) are shown.
2.4.3 Place and temporal information

The goal of a pitch perception model is to predict the pitch that a human listener would perceive in response to a given sound. Modern models can be split into two main components, as shown in Figure 2.13. The first block is an auditory model, as just described. The pitch processor then models the way the brain analyses the information carried on the auditory nerve to create a sensation of pitch. While there is general consensus regarding the principles of the auditory model, there is much more debate over the pitch processor. This central processor may combine information from both ears (Houtsma and Goldstein 1972), but a simpler monaural model is often assumed. The information carried by the auditory nerve is often described as being of two different kinds: place information and temporal information.

![Figure 2.13. Pitch perception model](image)

A simple measure of place information is the average neural firing rate at each place. Figure 2.14 shows the mean activity according to the Lyon model for the three example tones from §2.4.2. Each tone has a pitch of 200 Hz.

For a pure tone, the pitch corresponds to the place that has the maximum activity (Figure 2.14a). However, this is inadequate as a general model for pitch perception. The smallest perceptible difference in frequency between two pure tones (the frequency difference limen, FDL) is only around 3 Hz at 1000 Hz. It is difficult to explain such fine discrimination given the bandwidth of the auditory filters (Moore 1973; 1997). Furthermore, at high sound levels, the firing rate saturates across a wide region of the cochlear.

A greater difficulty occurs for harmonic tones. The excitation pattern has one peak for each resolved harmonic (Figure 2.14b). The amplitudes of the peaks naturally depend on the amplitudes of the harmonics, so the fundamental does not always produce the largest peak; in fact the pitch is unchanged even if the fundamental is removed entirely. Pattern recognition models were developed that seek to find the lowest common sub-harmonic, but even they fail to account for the tone with no resolved harmonics (Figure 2.14c), which still has a pitch of 200 Hz.
It appears that the pitch processor must make use of temporal information (i.e. firing times). Tones that have no resolved harmonics allow melody recognition (Moore and Rosen 1979) and judgement of musical intervals (§2.7.1) (Houtsma and Smurzynski 1990). As was seen in Figure 2.10c and Figure 2.11c, temporal information concerning the fundamental frequency is contained in the neural response. Burns and Viemeister (1976; 1981) convincingly established (in the face of some scepticism) that sinusoidally-amplitude-modulated noise had a pitch corresponding to the modulation frequency. The pitch was rather weak, but subjects were able to recognise melodies and musical intervals for modulation frequencies up to about 800 Hz. Again, the modulation frequency is reflected in the neural firing times.

For a low-frequency pure tone, the firing is phase-locked to the tone (Figure 2.10a). Phase-locking degrades for pure tones above 1000 Hz, and is absent at 5000 Hz (Moore 1997, p 185). This is consistent with the observation that pure tones above 4000 Hz do not evoke a clear sense of pitch; at these high frequencies, subjects are unable to recognise musical intervals (Ward 1954) or melodies (Attneave and Olson 1971).

### 2.4.4 Effect of harmonic content

Pitch conveyed by purely temporal means is not very clear. Houtsma and Smurzynski (1990)
found that musical interval judgements improved as more resolved harmonics were added to the tones. A variety of experiments have shown that pitch is dominated by the lower, resolved harmonics (reviewed in Plack and Oxenham 2005b).

Many studies have demonstrated better frequency discrimination with harmonic tones than with pure tones (e.g. Zeitlin 1964; Henning and Grosberg 1968; Spiegel and Watson 1984; Platt and Racine 1985). Moore and Peters (1992) measured frequency difference limens for pure tones and for harmonic tones in the same subjects. A subset of their data is replotted in Figure 2.15. The left panel shows the data for the young (aged 24 – 34) normal-hearing group, and the right panel shows the data for the elderly (aged 62 – 83) group, who had near-normal auditory thresholds below 2000 Hz. The dotted horizontal line indicates a one semitone frequency difference (about 6%) for reference (see §2.7.1). The elderly group had worse absolute performance than the young group, but showed similar trends for the effect of harmonic content. For the low fundamental frequencies shown in Figure 2.15, the frequency difference limen was highest for pure tones (labelled as harmonic 1), intermediate for harmonic tones comprising harmonics 1 – 5, and lowest for harmonic tones comprising harmonics 1 – 12. In other words, performance improved as more harmonics were added. These results are consistent with a central pitch processor that optimally combines information from multiple harmonics (Goldstein 1973; Wakefield and Nelson 1985).

The biological mechanism that utilises temporal information is not yet known. The following sections describe several proposed models.
2.4.5 Temporal pitch models

The auto-correlation model was proposed by Licklider (1951). An implementation described by Slaney and Lyon (1990; 1993) uses the Lyon cochlear model (Figure 2.7), which produces a multiple-channel output. Each cochlear channel signal $u_k(n)$ represents the instantaneous firing probability of an auditory neuron at place index $k$ and time index $n$. The short-term auto-correlation function is taken of each channel:

$$r(k, n, \tau) = \sum_{m=0}^{M-1} w(m) u_k(n - m) u_k(n - m - \tau)$$

(2.7)

where $w(m)$ is a window function, and $\tau$ is the lag. The auto-correlation of a periodic signal has peaks at lags that are multiples of the period. Slaney and Lyon called the collection of auto-correlations a correlogram, and displayed it as a movie. At each time index $n$, a new image $r(k, \tau)$ is produced, with place $k$ as the vertical axis, and lag $\tau$ as the horizontal axis. For a steady tone, the image is stable. The next step is to sum across place, yielding a summary auto-correlation function (SACF), a function of lag $\tau$. Finally, the lag $\tau_p$ that has the largest peak corresponds to the pitch period (i.e. the pitch frequency is $1/\tau_p$).

The auto-correlation model is consistent with a wide range of observed pitch phenomena involving both resolved and unresolved harmonics (Meddis and Hewitt 1991; Meddis and O'Mard 1997). However, it is a functional, rather than physiological model, as it performs processing operations on signals that represent instantaneous firing probability.

Licklider (1951) also proposed an equivalent neural mechanism (Figure 2.16). Each auditory neuron $A$ has a neuronal delay line $B_1$, $B_2$, $B_3$…; the delayed and undelayed spikes are fed to coincidence detectors $C_1$, $C_2$, $C_3$…; and finally integrators $D_1$, $D_2$, $D_3$… count the coincidences over short time windows, also receiving inputs from other neuronal channels at the same place. Lyon (1984) implemented this mechanism, calling it an auto-coincidence model. The neurons used a stochastic leaky integrate-to-threshold model and produced a 1-bit output at a 20 kHz sampling rate. There were 24 neurons for each cochlear channel, each with randomised model parameters. The coincidence detectors were AND gates and the integrators were counters. Because time intervals are measured between one spike and all the following spikes (up to the delay line duration), this is sometimes referred to as an all-order inter-spike intervals model.
An alternative pitch processor only examines the time intervals between one spike and the next, referred to as \textit{first-order} inter-spike intervals (Moore 1997). A subsequent stage searches for common intervals across channels. The time interval which occurs most often corresponds to the pitch period.

The auto-correlation pitch model sums auto-correlations from each cochlear channel uniformly across place. However, the possible temporal firing patterns of a particular neuron are partly constrained by its characteristic frequency. Consider for example the response of a neuron originating at the 500 Hz place to a variety of harmonic tones. For F0 = 500 Hz, there is a strong response and the inter-spike intervals are multiples of 2 ms. For F0 = 400 Hz, there is a smaller response to the fundamental, and the intervals are multiples of 2.5 ms. For F0 = 250 Hz, there is a large response to the second harmonic, with multiples of 2 ms again. For a much lower fundamental, such as F0 = 50 Hz, the tenth harmonic is not well resolved, so the response follows the envelope, with intervals clustered around 20 ms. Fundamental frequencies higher than 500 Hz produce little response, due to the sharp high-pass roll-off of the basilar membrane tuning, so intervals smaller than 2 ms are rare. Place and temporal cues are thus inter-related. A temporal model that employs this place information would analyse the inter-spike intervals of a neuron having characteristic frequency $f$ over time intervals in the range of about $0.5/f$ to $15/f$ seconds (Moore 1997). Bernstein and Oxenham (2005) used this principle to modify the auto-correlation pitch model of Meddis and O'Mard (1997), and found that it gave a better prediction of F0 discrimination performance for bandpass-filtered harmonic tones.

\subsection*{2.4.6 Phase transition models}

One major criticism of the auto-correlation (or auto-coincidence) model is that a calibrated neural delay line of up to 30 ms for each auditory neuron is not biologically plausible. Loeb, White \textit{et al.} (1983) proposed a spatial cross-correlation model, where the basilar membrane travelling wave delay fulfils an analogous function. In normal hearing, the response to a
resolved frequency component shows a large phase shift across a short distance of the basilar membrane at the place that has the corresponding characteristic frequency (§2.4.2, Figure 2.11a). A cross-correlation between nearby cochlear channels has similar properties to an auto-correlation. In this model, the central pitch processor compares neural firing times between neurons from closely-spaced points on the cochlea. A detector for a specific frequency looks for coincidences between firings on neurons that are one wavelength apart in place.

Van Schaik (2001) implemented an analog integrated circuit using this principle, as shown in Figure 2.17. The cochlear filterbank used a cascade of 104 second order filter sections. One periodicity detector was driven by two taps of the cascade, four filter sections apart. Each tap had an inner hair cell (IHC) model driving a neuron circuit to generate spikes. The coincidence detector was a simple AND gate. The periodicity detector was highly frequency selective, and robust to changes in the audio input level or signal to noise ratio.

![Figure 2.17 Periodicity detector (Van Schaik 2001) using coincidence detection between nearby cochlear channels](image)

Shamma (1985b) proposed an alternative neural mechanism, lateral inhibitory networks (LIN). The equivalent signal processing operation is to take the signal representing the probability of firing of a cochlear channel, and subtract a weighted sum of its neighbours, with weights decreasing with spatial separation. The simplest case is to take each channel and subtract its more basal neighbour. Figure 2.18 shows the result of applying this simple LIN to the 200 Hz tones of Figure 2.11. The phase transitions of the pure tone and resolved harmonics yield a "sharpened" response, which is even more localised in place. Unresolved harmonics create an excitation pattern where nearby channels are aligned in phase, which is suppressed by the LIN.
The underlying principle is that the phase characteristics of neighbouring auditory filters allow an individual resolved frequency component to be detected by its distinctive spatial pattern in a local region of the cochlea. It seems feasible that both excitatory and inhibitory mechanisms could be involved in this local detection process. A remaining issue is that because each harmonic is detected independently, there must be a more central processor which combines information from all the detected harmonics to find the common fundamental frequency.

2.5 Timbre perception

Timbre is the perceptual quality that distinguishes between two tones that have the same pitch, loudness and duration (Plomp 1976, chapter 6; Moore 1997, chapter 7). It allows the musical instrument that played a particular note to be identified. Timbre depends on many physical
properties of a tone. The most prominent is the spectral profile, i.e. the amplitudes of the harmonics ($a_n$ in Equation 2.1). Tones having strong high harmonics sound brighter.

In an admirable study (considering the primitive equipment of the time), Lichte (1941) investigated the brightness attribute of timbre. He generated a set of tones that had equal fundamental frequency and duration. Each tone comprised the first 16 harmonics, with amplitudes (in dB) varying linearly with harmonic number. The spectra of five such stimuli are shown in Figure 2.19; the dullest (bottom panel) had a spectral profile with a slope of -1 dB per harmonic; the brightest (top panel) had a slope of +1 dB per harmonic. The overall amplitude of each tone was firstly adjusted to equalise loudness. Pairs of stimuli were then presented and subjects were asked to rank them in brightness. Lichte constructed a brightness scale by converting percent correct scores into perceptual distance, and found a linear dependence of brightness on spectral slope. After further experiments with different spectral profiles, he concluded that "brightness is a function of the location on the frequency continuum of the midpoint of the energy distribution."

![Figure 2.19. Five of the harmonic tones used by Lichte (1941)](image_url)

Another aspect of timbre is that a musical instrument produces tones with a characteristic temporal envelope (in contrast to the idealised, infinite-duration harmonic tone described by
Equation 2.1). For example, a piano note has a rapid onset (attack), and a more gradual decay.

Being a multi-dimensional quality, timbre is well-suited to multi-dimensional scaling (MDS) studies (§2.3.5). Grey (1977) synthesised a set of 16 tones, equalized for pitch, loudness, and duration. The amplitude and frequency of each sinusoidal component was individually controlled by functions derived from the analysis of recordings of real musical instruments. A three-dimensional space provided a good fit to the MDS dissimilarity matrix. The first dimension was related to the spectral energy distribution; the second to the time alignment of the onsets and offsets of the harmonics; and the third to the amount of low-amplitude, high-frequency inharmonic components in the initial attack segment.

### 2.5.1 Independence of pitch and brightness

Plomp and Steeneken (1971; also reported in Plomp 1976, chapter 6) performed a multi-dimensional scaling study (§2.3.5) investigating the relationship between pitch and brightness. The nine stimuli were harmonic tones, with fundamental frequencies of 200, 250, or 315 Hz, having amplitude spectra centred at 2000, 2500, or 3150 Hz, as illustrated in Figure 2.20 (top panel). The dissimilarity matrix was well fit by a two-dimensional space (Figure 2.20, bottom panel), and the two perceptual dimensions were readily identified as pitch (corresponding to the physical parameter of fundamental frequency) and brightness (corresponding to spectral centroid frequency). In other words, pitch and brightness formed two independent perceptual dimensions. Furthermore, tones of differing pitch having the same spectral profile (aligned vertically in both panels of Figure 2.20) sounded more similar to each other than tones of differing pitch whose corresponding harmonics had the same amplitude (e.g. the diagonal of Figure 2.20 having \( F_c = 10 F_0 \)). This is further evidence that brightness depends on the spectral centroid.

Pitt and Crowder (1992, experiment 2a) presented pairs of synthesised harmonic tones to 50 subjects and asked them whether the pitch was the same or different. The set of six tones covered three fundamental frequencies (F4, G4, A4), and two types of harmonic content (either harmonics 1, 2, 3, 4, or harmonics 1, 5, 6, 7; all with equal amplitude). In the subset of trials where the two tones had the same harmonic content, the group mean score for the same/different task was over 90%. However, when the fundamental frequency was the same, but the harmonic content differed, performance was highly correlated with musical training. The subjects categorised as musicians scored at or near 100%. In contrast, 15 of the subjects categorised as non-musicians scored 0%, i.e. they consistently responded that the pitch had changed when the fundamental frequency was constant but the amplitude spectrum changed. Pitt and Crowder concluded that "people differ greatly in their ability to disregard changes in timbre while attending to pitch". An alternative interpretation of the results is simply that non-musicians had a less-strict definition of the term "pitch", and were unfamiliar with the
distinction between pitch and brightness.

Figure 2.20. Pitch and brightness MDS study (Plomp 1976, chapter 6).

Top panel: Stimulus set. The spectral profile is shown as a dotted line.

Bottom panel: Resulting perceptual space.

Labels indicate centre frequency (Fc) and fundamental frequency (F0), in Hertz.

2.6 Vowel perception

Speech sounds can be categorised as voiced or unvoiced. Voiced sounds are produced by vibrations of the vocal folds in the larynx. Vowels are voiced sounds, with waveforms that are relatively stable and nearly periodic, resembling harmonic tones. Their amplitude spectrum is
determined by the shape of the vocal tract, i.e. the position of the jaw, lips and tongue. The vocal tract acts as a filter which accentuates the harmonics that lie in particular frequency ranges. The resulting peaks in the spectral profile are called *formants*, and are numbered F1, F2, F3, etc, starting at the lowest frequency. Different vowels can be distinguished by the frequencies of their first and second formants (F1 and F2). Vowel perception is closely related to timbre perception, both involving an analysis of the spectral profile. Vowel identity is independent of voice pitch: the melody of a song is unaffected by the lyrics. This is analogous to the perceptual independence of brightness and pitch. An experiment using sung vowel stimuli will be reported in §10.2.4.

It is possible to attend to one person speaking in the presence of another competing speaker. This segregation of one speaker from the other can be done monaurally, despite the two voices having spectra that completely overlap in frequency. One of the cues used is voice pitch. Assmann and Summerfield (1990) presented two simultaneous synthetic vowels and asked subjects to identify both vowels. Performance improved substantially as the F0 difference increased from zero to four semitones. Assmann and Summerfield applied the auto-correlation pitch model (§2.4.5) to each stimulus. They used a 256-channel filterbank, and estimated the two F0s by finding the two largest peaks in the summary auto-correlation function. They then sampled each of the 256 auto-correlation channels at the two corresponding lags. The auto-correlation samples, as a function of place, formed a "synchrony spectrum" for each pitch estimate. Finally, the two vowels were identified by comparing the two synchrony spectra against the formant pattern of each candidate vowel. This model was a good predictor of the performance of the subjects as a function of F0 difference.

### 2.7 Music perception

#### 2.7.1 Musical intervals and scales

Two simultaneous pure tones, with frequencies differing by less than an auditory filter bandwidth, produce a sensation of "roughness" or *dissonance*. This is presumably due to the resulting amplitude modulation (beating) at the difference frequency.

Two simultaneous harmonic tones produce a pleasant sensation of *consonance* when there is a small integer ratio between the two fundamental frequencies. These relationships are referred to as musical intervals. The most consonant frequency ratio of 2:1 is called an *octave*. Two notes an octave apart sound very similar, and are given the same note name in Western music notation (e.g. C). The next-most consonant frequency ratio of 3:2 is called a *fifth*. Plomp and Levelt (1965) demonstrated that the consonance or dissonance of a musical interval can be explained by examining the frequencies of the two sets of harmonics. Smaller integer ratios result in more
harmonics coinciding in frequency, producing greater consonance.

In the *equal-temperament* tuning, there are 12 equally-spaced notes in an octave, as listed in Table 2.2. The frequency ratio between adjacent notes is thus \(2^{1/12}\) (about a 6% change), and is called a *semitone*. This tuning yields intervals between notes that are fairly close to small integer ratios. The notation used in this thesis appends a number representing the octave to the note name, e.g. middle C is C4. Frequencies are calculated relative to a reference of 440 Hz for A4. The final column of Table 2.2 lists the note frequencies for octave 4, which starts with C4.

<table>
<thead>
<tr>
<th>Note name</th>
<th>Number of semitones above C</th>
<th>Musical interval from C</th>
<th>Frequency ratio (approximate)</th>
<th>Frequency (Hertz) for octave 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>12</td>
<td>octave</td>
<td>1:2</td>
<td>523 (C5)</td>
</tr>
<tr>
<td>B</td>
<td>11</td>
<td>major seventh</td>
<td>8:15</td>
<td>494</td>
</tr>
<tr>
<td>A#</td>
<td>10</td>
<td></td>
<td></td>
<td>466</td>
</tr>
<tr>
<td>A</td>
<td>9</td>
<td>major sixth</td>
<td>3:5</td>
<td>440</td>
</tr>
<tr>
<td>G#</td>
<td>8</td>
<td></td>
<td></td>
<td>415</td>
</tr>
<tr>
<td>G</td>
<td>7</td>
<td>perfect fifth</td>
<td>2:3</td>
<td>392</td>
</tr>
<tr>
<td>F#</td>
<td>6</td>
<td></td>
<td></td>
<td>370</td>
</tr>
<tr>
<td>F</td>
<td>5</td>
<td>perfect fourth</td>
<td>3:4</td>
<td>349</td>
</tr>
<tr>
<td>E</td>
<td>4</td>
<td>major third</td>
<td>4:5</td>
<td>330</td>
</tr>
<tr>
<td>D#</td>
<td>3</td>
<td></td>
<td></td>
<td>311</td>
</tr>
<tr>
<td>D</td>
<td>2</td>
<td>major second</td>
<td>8:9</td>
<td>294</td>
</tr>
<tr>
<td>C#</td>
<td>1</td>
<td></td>
<td></td>
<td>277</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>unison</td>
<td>1:1</td>
<td>262 (C4)</td>
</tr>
</tbody>
</table>

Table 2.2 Note names, intervals and frequencies of the equal-tempered scale

### 2.7.2 Melody perception

A melody is a sequence of notes that make up a musical phrase. Melodies in this thesis will be represented by plots such as Figure 2.21, which shows the first few bars of Old MacDonald. Each note is shown as a horizontal line, with length indicating duration (in arbitrary units referred to as beats), and vertical position indicating pitch. The vertical axis is linear in semitones (i.e. logarithmic in frequency), unlike a traditional musical staff, which has non-uniform semitone spacing. The timing of a melody (*tempo*) is described by specifying how many beats occur in a minute.
To transpose a melody means to add (or subtract) a constant number of semitones to each note, i.e. to shift it up or down in Figure 2.21. This does not change the melody; only the relative changes in pitch are important. For example, Old MacDonald is characterised by the intervals between the notes shown in red in Figure 2.22. The contour of the melody is defined as the signs of the intervals, shown in the upper part of Figure 2.22, i.e. the directions (up or down) of the pitch steps, disregarding the sizes of the intervals (Dowling and Fujitani 1971).

Dowling and Fujitani (1971) investigated the roles of interval and contour in melody recognition. In each trial, the subject heard a pair of melodies. The first melody (the reference
R) was a new randomly generated five-note sequence. The second melody was either a transposition of the reference (T) or else a variant sequence D that had the same starting note as T but a different contour. Examples are shown in Figure 2.23, with the intervals indicated. Subjects were asked whether the melodies (ignoring the transposition) were the same or different, i.e. "same" trials were RT, and "different" trials were RD. Performance was high, indicating good discrimination of T and D. In a second condition, the variant C had the same contour as the reference R but different interval sizes (Figure 2.23), i.e. "same" trials were RT, and "different" trials were RC. Performance was near chance, indicating poor discrimination of T and C. Dowling and Fujitani concluded that when a subject was presented with an unfamiliar melody, the contour was stored in their short-term memory, but not the interval sizes.

Dowling and Fujitani (1971) also tested familiar melody identification (§5.2.3) in 28 normal hearing subjects, and obtained a mean score of 99%. In a second condition, subjects were presented with distorted versions of the melodies, where notes were shifted in pitch while preserving the contour, and the mean identification score was substantially worse (59%). Dowling and Fujitani concluded that both the contour and the exact interval sizes of a familiar melody were stored in long-term memory.

Figure 2.23 Example of melodies used in Dowling and Fujitani (1971)
3 Cochlear implants

3.1 Introduction

This chapter briefly outlines some of the clinical aspects of cochlear implants, describes the implant system hardware and the principles of operation. It then summarises cochlear implant psychophysics, i.e. the perception of electrical stimulation.

3.2 Clinical aspects

Many people with hearing loss can obtain benefit from a hearing aid. However, if the inner hair cells of the cochlea are not functioning, then amplification of sound will not help, and such people are candidates for cochlear implants.

Deaf children are routinely implanted at less than two years of age, as it has been found that younger children achieve better outcomes. Cochlear implants are generally not recommended for pre-lingually deaf adults (adults whose lack of hearing prevented them from learning to speak).

Post-lingual deafness refers to hearing loss later in life. Hearing can be damaged by illness, certain drugs, trauma to the ear, or long-term exposure to loud sound. Hearing loss is sometimes progressive, and often affects high frequencies first. Profound deafness has a severe impact on education and employment and can lead to a sense of social isolation. Post-lingually deaf adults who receive cochlear implants usually obtain good speech perception, and many can use the telephone. There is a trend for poorer outcomes for recipients who had a long duration of deafness before receiving their implant; this is thought to be due to a progressive decline in the numbers and condition of the spiral ganglion cells in a deaf ear.

3.3 Cochlear implant systems

All commercially-available cochlear implant systems consist of two components: a surgically implanted stimulator, and an external processor. Early processors were body-worn, but more recent processors are worn behind the ear. The battery-powered processor contains one or two microphones, electronic circuitry for processing the sound, and user controls. It has a radio frequency (RF) coil for sending power and commands to the stimulator, and receiving status and telemetry data from it. The stimulator extracts power and data from the received RF signal and applies pulses of current to an array of electrodes inserted into the scala tympani of the cochlea. These pulses stimulate the spiral ganglion cells, giving the perception of sound. Having multiple electrodes allows the tonotopic response of a normal cochlea to be emulated: electrodes
positioned at the basal end of the cochlea are activated in response to high frequency sounds, and electrodes at the apical end are activated by low frequencies. A detailed description of the sound processing is given in Chapter 4.

The three major manufacturers of cochlear implants are Cochlear Ltd (Australia), Advanced Bionics Corporation (USA), and MED-EL (Austria).

There have been three generations of Nucleus cochlear implant from Cochlear Ltd. They can be distinguished by the integrated circuit (IC) that implements their functions, as shown in Table 3.1. All of the research subjects in this thesis had either a Nucleus 24 or Nucleus Freedom implant (details are given in §15). The Nucleus Freedom implant and processor are shown in Figure 3.1.

<table>
<thead>
<tr>
<th>System name</th>
<th>Implant model</th>
<th>IC</th>
<th>RF link frequency (MHz)</th>
<th>Total pulse rate (pps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nucleus 22</td>
<td>CI22M</td>
<td>CIC1</td>
<td>2.5</td>
<td>1500</td>
</tr>
<tr>
<td>Nucleus 24</td>
<td>CI24M CI24R</td>
<td>CIC3</td>
<td>5.0</td>
<td>14400</td>
</tr>
<tr>
<td>Nucleus Freedom</td>
<td>CI24RE</td>
<td>CIC4</td>
<td>5.0</td>
<td>31500</td>
</tr>
</tbody>
</table>

Table 3.1 Nucleus cochlear implant models

Figure 3.1 Nucleus Freedom cochlear implant system.
CI24RE stimulator (left) and Freedom processor (right)

### 3.4 Electrical stimulation

The term *electrode* refers to the physical electrode contact. Nucleus implants have 22 intracochlear electrodes, numbered from E1 at the basal end to E22 at the apical end. In addition, the Nucleus 24 and Freedom implants have two extra-cochlear electrodes. The first (ECE1) is a ball electrode connected by a separate lead wire, visible in Figure 3.1. The second (ECE2) is a platinum plate mounted on (but insulated from) the titanium package of the implant.

To create a path for current to flow, at least two electrodes must be involved (sometimes
referred to as the active electrode and the reference electrode). The stimulation mode describes the configuration of the electrodes. In bipolar mode, current flows between two intra-cochlear electrodes. In the Nucleus 22 system, the abbreviation BP indicates the use of two adjacent intra-cochlear electrodes (e.g. E4 – E5); BP+1 indicates an additional spacing of one electrode (e.g. E4 – E6); BP+2 would use E4 – E7, and so on. In common ground (CG) mode, one intra-cochlear electrode is used as the active electrode, and all the remaining intra-cochlear electrodes connected together are used as the reference. The most commonly used mode in modern systems is monopolar mode, in which current flows between an intra-cochlear electrode (active) and an extra-cochlear electrode (reference). The corresponding abbreviations for Nucleus systems are: MP1 uses ECE1 as the reference electrode; MP2 uses ECE2; and MP1+2 uses ECE1 and ECE2 connected together. The default is MP1+2, because it gives the lowest impedance, and so requires the least voltage to deliver a specified current. The Nucleus Freedom implant (CI24RE) has the ability to connect two adjacent intra-cochlear electrodes together to form the active electrode, with one or both extra-cochlear electrodes used as the reference; this is known as monopolar double-electrode mode.

Because of the variety of possible stimulation modes, it is sometimes helpful to make a distinction between an electrode and a stimulation channel. A stimulation channel (or sometimes simply channel) is one specific electrode configuration used to deliver a pulse (e.g. E7 in MP1+2 mode). Each stimulation channel defines a stimulation site in the cochlea. With 22 intra-cochlear electrodes, there are 22 possible monopolar channels; but only 21 possible monopolar double-electrode channels (since the last electrode has no "mate"). Similarly there are only 21 BP channels, 20 BP+1 channels, and so on. For monopolar mode there is a one-to-one correspondence between an intra-cochlear electrode and a stimulation channel, and so the two terms are often used interchangeably.

Current in the Nucleus 24 system is specified by an 8-bit value. The current \( i \) in microamps is related to the current level \( c \) in clinical units, by:

\[
i = i_0 \left( \frac{i_m}{i_0} \right)^{\frac{c}{c_m}}
\]

(3.1)

where \( i_0 = 10 \, \mu A \) (the minimum current, produced by \( c = 0 \)), and \( i_m = 1750 \, \mu A \) (the maximum current, produced for \( c = c_m = 255 \)). This can also be expressed as an exponential relationship:

\[
i = i_0 e^{rc}
\]

(3.2)

where \( r = 0.0203 \). Each step is about a 2% increase in current (0.176 dB).

The pulse waveform used in Nucleus implant systems is shown in Figure 3.2. This biphasic
pulse has two phases of equal duration and amplitude (but opposite polarity). Current flows from the reference electrode to the active electrode in phase one, and in the reverse direction in phase two, so that there is no net direct current flow. In every experiment in this thesis, the subjects were stimulated in monopolar mode (usually MP1+2) with pulses having 25 µs phase width and 8 µs phase gap.

Nucleus implants have a single current source, and use electronic switches to route the current to the required electrodes. Thus only one channel is stimulated at a time; this is known as sequential or interleaved stimulation. An example of such a stimulation pattern, or pulse sequence, is shown in Figure 3.3. Typical stimulation rates on each electrode range from 250 to 3500 pulses per second (pps). The phase width is usually kept constant, so at longer time scales it can be convenient to represent each pulse by a single vertical line, with height representing amplitude, as shown in Figure 3.4.
3.5 Loudness

3.5.1 Single channels

Traditionally, the first step in fitting a cochlear implant is to measure the threshold (T level) and maximum comfortable level (C level) for each stimulation channel. Short bursts of pulses are presented on a single channel (using the stimulation mode, pulse rate and pulse width that is to be used in the subject's processor) at increasing current levels. The popularity of monopolar mode is due to the fact that T and C levels are lower in monopolar mode than in bipolar mode. Note that, strictly speaking, T and C levels are associated with stimulation channels, not electrodes. The electrical dynamic range is defined as the ratio between C and T currents (each measured in microamps), or the difference between C and T levels (each measured in clinical units).

Fu and Shannon (1998) obtained numerical loudness estimates (§2.3.4) from three Nucleus 22 recipients for 500 pps pulse trains, as the current was varied between the T and C level. Consistent with Stevens' power law (Equation 2.6), the loudness estimates $L$ were well-fit by a power function of current $i$ in microamps, with mean exponent $\beta = 2.7$:

$$L = k i^{\beta}$$ (3.3)

Equation 3.3 can be rewritten in terms of a reference current $i_{\text{ref}}$ that produces a loudness of 1.0:
Substituting Equation 3.2, loudness can be expressed as a function of current level \( c \) in clinical units, where current level \( c_{\text{ref}} \) produces a loudness of 1.0:

\[
L = \left( \frac{i}{i_{\text{ref}}} \right)^{\beta}
\]  

(3.4)

Loudness is often expressed on a logarithmic scale. The power law predicts that log loudness is a linear function of current level \( c \) in clinical units, with slope \( a = \beta r \):

\[
\log(L) = a \left( c - c_{\text{ref}} \right)
\]  

(3.6)

3.5.2 Multiple channels

When multiple channels are stimulated sequentially, each channel contributes to the loudness of the stimulus. McKay, Remine et al. (2001) asked recipients to adjust the current of a two-channel pulse sequence so that it had the same loudness as a reference single-channel pulse sequence. They found that varying the spatial separation of the two channels had a much smaller effect than varying pulse rate or the overall loudness level. A method for predicting the loudness of multiple-channel stimuli was subsequently developed by McKay et al. (2003). It relied on the approximation that the amount of loudness summation was independent of the channel spacing. Each pulse contributed to loudness independently, regardless of the channel it was presented on. The model firstly calculated the loudness contribution of each pulse, then summed the loudness contributions from all pulses that fell in a 2 ms time window. The model gave good predictions of the current levels needed to balance the loudness of a variety of stimuli.

In estimating the loudness contributions of individual pulses, McKay et al. (2003) observed that log loudness was a linear function of current level at low current levels (Equation 3.6), but at higher current levels, loudness grew much more steeply. For six Nucleus 24 recipients, the knee-point current at which the growth changed was close to the corresponding C-level for a 500 pps pulse train.

3.6 Place pitch

Early psychophysical studies with multiple-channel cochlear implants (Simmons et al. 1965; Eddington et al. 1978; Tong et al. 1982; Shannon 1983; Townshend et al. 1987) showed that the percepts produced by stimulating individual electrodes could be ranked in a generally tonotopic order, corresponding to the location of the intracochlear electrodes. The term place pitch is commonly used to describe this percept. Apical electrodes have lower place pitch than
Nelson et al. (1995) undertook a comprehensive study of place pitch with fourteen Nucleus 22 recipients. All possible pairs of usable electrodes were pitch-ranked, and scores were converted to $d'$ per millimetre of distance along the electrode array. The place pitch sensitivity varied widely across subjects: some subjects obtained perfect scores for adjacent electrodes, while others needed a separation of up to 18 electrodes (13.5 mm). Several subjects had regions of very poor sensitivity to place pitch, and there were instances of reversals in the rankings of adjacent electrodes (i.e. the more apical electrode was ranked higher in pitch). If place-pitch represents a single perceptual dimension, then $d'$ sensitivity should be additive; e.g. the sum of $d'$ for ranking electrodes E1 – E2 and $d'$ for ranking electrodes E2 – E3 should equal $d'$ for ranking electrodes E1 – E3. This hypothesis could not be tested for the best-performing subjects (because they had perfect scores for adjacent electrodes), but for the poorer-performing subjects it was found that $d'$ was additive.

3.6.1 Intermediate place pitch

Townshend et al (1987) found that when two electrodes were simultaneously stimulated, the pitch varied smoothly between the pitch of the two individual electrodes as the ratio of the currents was varied. Interest in this current-steering effect has recently been revived. Donaldson et al (2005) obtained pitch-ranking results from six Clarion CII cochlear implant recipients. One stimulus was a sequence of pulses on a single electrode; the other was a sequence of pulses on two adjacent electrodes simultaneously, with the proportion of current on the more basal electrode denoted $\alpha$. The method of constant stimuli (§2.3.3) was used to obtain a psychometric function of sensitivity ($d'$) as a function of $\alpha$. The experiment was repeated using a 3-down, 1-up adaptive rule (§2.3.3), and the estimated just-noticeable difference ($d' = 1.16$) agreed well with that obtained by interpolation of the psychometric function. Consistent with the results of Nelson et al. (1995) described above, there was substantial variation in place-pitch sensitivity across subjects and electrodes: the lowest threshold was $\alpha = 0.11$, whilst one subject could not discriminate one pair of adjacent electrodes. Sensitivity was better for medium-loud stimuli than medium-soft. This current steering effect has been used to attempt to enhance place-pitch cues in processing strategies (§6.5.1 and §6.5.2).

The Nucleus Freedom implant (CI24RE) has the ability to connect adjacent electrodes together to create a "double electrode", which can be stimulated in monopolar mode. Busby and Plant (2005) measured place-pitch ranking with eight CI24RE recipients and found that most subjects ranked the place pitch percept of the double electrode intermediate to that of the two single electrodes.

McDermott and McKay (1994) demonstrated that intermediate place-pitch percepts could also
be created by sequential stimulation of nearby channels. The time between pulses was 400 μs, which is less than the neural refractory period. They suggested that the effect was caused by spatial overlap of the two populations of neurons excited by each channel, and that the place-pitch percept could be determined by the centroid of the combined spatial distribution. They pointed out that this effect was likely to occur in the normal operation of the SPEAK strategy (§4.6), where groups of nearby channels were activated in response to a spectral peak. Because the filter bands overlap, small changes in formant frequency produced variation in the current levels on the activated channels. It was a plausible explanation for the improvement in vowel perception obtained with SPEAK, compared to earlier strategies that activated a single channel for each formant. This hypothesis will be tested experimentally in §8.4.

A subsequent multi-dimensional scaling study (McKay et al. 1996) showed that the percepts evoked by the dual-channel stimuli (with minimum inter-pulse delays) lay at intermediate points along a single-dimensional arc connecting the percepts evoked by the individual channels. Pitch-ranking sensitivity (d') for intermediate place-pitch percepts with sequential stimulation was measured by Kwon and van den Honert (2006). The ratio of the currents on the two channels was varied, and pitch ranking performance was comparable to that reported for simultaneous stimulation by Donaldson et al. (2005). In both cases, there was significant variation between recipients, and across the electrode array. A similar experiment will be described in §8.3.

### 3.6.2 Centroid model of place pitch

A quantitative model of place-pitch perception was proposed by Laneau et al. (Laneau et al. 2004; Laneau 2005, chapter 5). In that study, four cochlear implant recipients pitch-ranked harmonic tones processed by the ACE strategy (§4.6), using standard and alternative filterbanks. In one condition, the filter envelopes were low-pass filtered to remove temporal cues, leaving only place cues. Laneau et al. hypothesised that the place pitch of each stimulus was determined by the centroid \( c \) (or "centre of gravity") of the stimulation pattern, calculated as:

\[
c = \frac{\sum_{k} k a(k)}{\sum_{k} a(k)}
\]  

(3.7)

where \( k \) is the channel number, and \( a(k) \) is the amplitude of the corresponding filter envelope. The ability to discriminate between two stimuli on the basis of place pitch should depend on the distance between the two centroids; i.e. the sensitivity index \( d' \) for a pitch-ranking task using place-pitch cues alone should be given by:
where \( d' \) denotes the difference between the centroids of the two stimuli, and \( m \) is a constant for each subject characterising their usage of the place-pitch cue. Thus, from Equation 2.2 (see §2.3.1), the proportion-correct score \( p \) should be related to the centroid difference as:

\[
p = \Phi \left( \frac{m c'}{\sqrt{2}} \right)
\]

where \( \Phi \) is the cumulative normal distribution. Laneau et al. found that the pitch-ranking scores, pooled across the four subjects, were modelled reasonably well by Equation 3.9. In another study (Laneau and Wouters 2004; Laneau 2005, chapter 3), the same four subjects pitch-ranked stimuli consisting of constant-amplitude pulse trains interleaved across multiple consecutive channels. In each trial, the two stimuli had the same number of active channels, but the set of channels was shifted either apically or basally by one or more electrode places. The place-pitch sensitivity was independent of the number of channels in the stimuli, i.e. increasing the amount of spatial overlap in the two stimulation patterns had no significant effect. These results were also consistent with the centroid model. Furthermore, the just noticeable differences calculated from each individual subject's scores were in good agreement between the two studies.

The centroid model is also consistent with the results of Cohen et al. (1996), where the centroid of the forward masking distribution corresponded to the numerical pitch estimate (§2.3.4).

The place-pitch centroid model is further explored in §8.5 and §10.3.3.

### 3.7 Temporal pitch

Varying the pulse rate on a single channel provides a rate pitch percept. The upper limit at which an increase in pulse rate is no longer perceived as an increase in pitch is typically about 300 pps (Simmons et al. 1965; Eddington et al. 1978; Tong and Clark 1985; Zeng 2002), but ranges up to 1000 pps for some individuals (Townshend et al. 1987). If the pulse width and current of each pulse are kept constant, then increasing pulse rate also increases loudness (Tong et al. 1983a; Busby and Clark 1997). Therefore the current is usually adjusted to balance the loudness of the stimuli before pitch comparisons are made.

Varying the frequency of amplitude modulation applied to a high-rate pulse train on a single channel also provides a pitch percept. On-off modulation was investigated in two Melbourne studies (Tong et al. 1983a; Busby and Clark 1997). The pulse pattern was referred to as a Multiple Pulse per Period (MPP) sequence, in contrast to the usual Single Pulse per Period (SPP) sequence (Figure 3.5). Because a fixed duty cycle of 50% was used, the total number of pulses per second in the MPP sequences did not change as the modulation frequency varied, so
it was hypothesised that the variation in loudness with modulation frequency would be minimal. This was true for the single subject in the early study (Tong et al. 1983a), but only for five out of 14 subjects in the later study (Busby and Clark 1997), with the remaining nine subjects showing substantial drops in loudness as the modulation frequency increased. Experimental work investigating the pitch percepts produced by MPP sequences will be reported in Chapter 9.

![Figure 3.5](image)

**Figure 3.5** Three types of pulse sequence, all having a period of 10 ms. (a) Single pulse per period (SPP), (b) Multiple pulse per period (MPP), (c) Sinusoidally amplitude-modulated (SAM)

Shannon (1992) investigated sinusoidal phase-width modulation of high-rate pulse trains on single channels. With a 1000 pps carrier rate, four Nucleus 22 recipients were most sensitive to modulation frequencies in the range 80 – 100 Hz. They required increasing modulation depths for detection as modulation frequency increased, and had limited ability to detect modulation frequencies above 300 Hz. A later study with seven Nucleus 22 subjects (Busby et al. 1993) replicated these findings. In that study, the four post-lingually deafened subjects performed better than the three pre-lingually deafened subjects, and the best performer (subject P4) had the shortest duration of deafness prior to implantation (four years). Busby et al. (1993) concluded that some of the large inter-subject variability in temporal processing ability was due to differences in neural survival.

The pitch of sinusoidally amplitude-modulated (SAM) pulse trains is of particular interest as this type of stimuli are produced by commonly-used sound processors in response to harmonic tones (§4.6). When the pulse rate is more than four times the modulating frequency, then the envelope of the pulse train is a good representation of the sinusoidal modulating waveform (e.g. Figure 3.5c). McKay and McDermott (1994) asked five Nucleus 22 subjects to rank the pitch of two SAM pulse trains with modulation frequencies of 150 and 200 Hz (a five semitone interval). When the carrier pulse rate was above 800 pps, their scores were similar to those obtained for pitch-ranking unmodulated pulse trains (SPP sequences) with pulse rates of 150 and 200 pps (with four of the subjects scoring at or near 100% for both conditions).
One musically trained recipient was extensively tested by McDermott and McKay (1997). His difference limens for the modulation frequency of SAM pulse trains were similar to those for pulse rate. His ability to produce and estimate musical intervals was similar for stimuli using either modulation frequency or pulse rate to convey pitch.

In summary, the perceptual effects of varying pulse rate and modulation frequency are similar, and both can be categorized as temporal pitch. As mentioned in §2.4.3, normal hearing listeners can derive purely temporal pitch from sinusoidally-amplitude-modulated noise for modulation frequencies up to at least 800 Hz (Burns and Viemeister 1976; 1981). The significantly lower temporal pitch limit of about 300 Hz for most cochlear implant recipients may be a consequence of deafness. Shepherd and Javel (1997) found that while neurons in a normal cat cochlea were capable of entrainment to electrical stimulation (i.e. responding to every stimulus pulse) at up to 600 – 800 pps, neurons in a cat deaf for four years were limited to rates below 400 pps. This was thought to be due to degeneration of the spiral ganglion cells.

3.7.1 Temporal pitch models

The pitch models described in §2.4.5 can be applied to predict the pitch evoked by simple electrical pulse trains. Consider a pulse sequence with a rate of 200 pps. Neural firing will be tightly synchronised to each stimulation pulse, giving 5 ms inter-spike intervals. An auto-correlation of the inter-spike intervals will have a strong peak at 5 ms lag, predicting a 200 Hz pitch. Because the inter-spike intervals are uniform, a model that only looked at first-order intervals would also predict a 200 Hz pitch.

Carlyon and colleagues conducted a series of experiments investigating purely temporal pitch on a single channel (Carlyon et al. 2002; van Wieringen et al. 2003; Carlyon et al. 2008). The reference stimulus was generated by taking an SPP sequence (Figure 3.5a) with a 5 ms period (i.e. 200 pps), and delaying every second pulse by 1 ms, so that the intervals between pulses alternated between 4 and 6 ms. Subjects performed a pitch-ranking task using the method of constant stimuli (§2.3.3). In each trial the reference "4 – 6" sequence was compared to an SPP sequence that had a period in the range of 3 to 12 ms. A psychometric function was fitted to the results to find the SPP period that gave the same pitch as the 4 – 6 sequence. The matching SPP period was in the range 5.3 to 6.2 ms (mean 5.7 ms) for five Laura implant recipients (Carlyon et al. 2002) and five Nucleus recipients (Carlyon et al. 2008). A similar experiment was also performed with normal hearing subjects. The stimuli were acoustic pulse trains (with the same timing as the electrical pulse sequences) which were band-pass filtered between 3900 and 5400 Hz to remove any resolved harmonics (§2.4.2). The mean matching period was 5.6 ms, very close to the mean cochlear implant result, suggesting that a common mechanism was involved for both acoustic and electric stimuli.
Carlyon et al. (2002) applied the Meddis and O'Mard (1997) autocorrelation model, and found that the summary autocorrelation function had its largest peak at 10 ms, with smaller peaks at 4 and 6 ms, none of which corresponded to the matching SPP period. Thus their results were not consistent with a pitch model that examines all-order interval statistics. Carlyon et al. argued that the results could be explained by a model that examined only first-order intervals. It would however have to perform an averaging operation on the intervals.

The issue of all-order versus first-order inter-spike intervals is taken up in §9.5.

3.8 Relationship between place and temporal cues

There are two schools of thought on the relationship between place cues and temporal cues in cochlear implants: that they are independent, or that they are interchangeable.

3.8.1 Place and temporal cues are independent

The first school of thought is exemplified by the following quotation from a pioneering Stanford University study (Simmons et al. 1965):

> We must be cautious in interpreting these observations on pitch since it appears that we are dealing with auditory perceptions containing multiple pitch components. It does seem clear enough that one pitch quality is characteristic of each electrode – thereby corresponding to a place-pitch representation – and that another quality is independent of electrode selection but is associated with stimulus repetition rate – corresponding to a volley-pitch representation.

It is clear that there are two independent attributes at the neural level: the electrode determines which nerves fire, and the pulse rate determines the firing times. It would be surprising if these two very different neural attributes produced the same sensation; instead the expectation is that place and rate should be qualitatively different.

The early University of Melbourne researchers preferred to use the term sharpness instead of place-pitch, reporting that subjects described the sensation as ranging from dull for the apical electrodes to sharp for the basal electrodes (Tong et al. 1982). In a multidimensional scaling (§2.3.5) study with the first Melbourne cochlear implant recipient (Tong et al. 1983a), there were nine stimuli. Each stimulus was a MPP pulse sequence (§3.7) at one of three different modulation rates, on one of three different channels. The physical parameters of the stimulus set can thus be represented in a two-dimensional grid, as shown in Figure 3.6a. The subject's resulting dissimilarity matrix was also well fit by a two-dimensional space, as shown in Figure 3.6b. Thus electrode place and modulation rate formed two independent perceptual dimensions. This was the rationale behind the early Nucleus stimulation strategies (§6.4.1), in which formant
frequency was represented by electrode position, whilst fundamental frequency was represented by pulse rate.

![Figure 3.6 Place and rate MDS experiment (Tong et al. 1983a)](image)

(a) Stimulus parameters (b) Perceptual space

In a discrimination task with small concurrent place and rate changes (where the subject was asked to detect a change in the stimulus), McKay et al. (2000) found results consistent with optimal processing of independent observations.

Pijl (1997b) investigated the effects of rate and place in a pitch-matching experiment with two Nucleus 22 cochlear implant recipients. In each trial, a reference and a comparison electrode were selected from a set of three electrodes (apical, mid, and basal). The pulse rate on the reference electrode was chosen from a set spanning the range 82 to 464 pps. The subject was asked to adjust the pulse rate on the comparison electrode until its pitch matched that of the reference. When the comparison electrode was the same as the reference, subjects matched the pulse rates very precisely, with a mean difference of only 0.4%. Crucially, when the comparison and reference were on different electrodes, the pulse rates were still adjusted to be almost equal, with a mean pulse rate difference of 4.4%. While this rate match was less precise than the same-electrode condition, it was still small in musical terms (0.74 semitones), and was insignificant compared to the place difference (the two electrodes were as much as 14 electrodes apart, corresponding to several octaves in characteristic frequency).

Even in normal hearing, pitch matching becomes more difficult in the presence of timbre differences; in a very similar task, Platt and Racine (1985) measured a mean error of about half a semitone for non-musicians when matching the pitch of a harmonic tone to a pure tone (and vice versa). The results of Pijl (1997b) support the view that the pitch of a low-rate pulse train is determined by the pulse rate; electrode position affects timbre, but does not substantially affect pitch.
3.8.2 Place and temporal cues are interchangeable

The second school of thought is that place cues and temporal cues are in some sense interchangeable. In an early study with a single subject, Eddington (1978; 1980) performed the obligatory measurements of electrode place pitch ranking and pulse rate difference limens. The subject was also asked to give numerical pitch estimates (§2.3.4) for a set of 18 single-channel stimuli. Each stimulus was a pulse train on one of the six electrodes at one of three rates (100, 175 and 300 pps). The results are reproduced in Figure 3.7. Eddington (1978) concluded that varying the pulse rate shifted the pitch over a range determined by the electrode position.

![Figure 3.7 Pitch scaling results (Eddington 1980, Figure 3)](image)

A more comprehensive series of pitch scaling studies by Fearn et al (Fearn et al. 1999; Fearn and Wolfe 2000; Fearn 2001, chapter 3) obtained similar results. For example, Figure 3.8 reproduces the average pitch estimates of four subjects for a stimulus set containing an apical, medial, and basal electrode, stimulated at rates ranging from 100 to 1000 pps. The pitch estimates depended strongly on both place and rate.
Another similar study was undertaken by Zeng (2002). Zeng explained that the autocorrelation model of pitch perception (§2.4.5) predicted that the pitch should depend only on the stimulation rate, and be independent of electrode position (as was found by Pijl (1997b)). He then claimed that this model was inconsistent with the joint dependence of pitch on rate and place that he observed. However, Zeng's own results (reproduced in Figure 3.9) did not justify this claim. The pitch estimates only had a consistent relationship with place for pulse rates above 300 pps, where rate cues were weak. At lower rates, where rate cues were strong, Zeng's data was equivocal: two out of the four subjects (AM and JM in Figure 3.9) provided lower pitch estimates for the basal channels than the apical channels near 100 pps (the opposite of that expected from the place cue), and there was no significant difference between pitch estimates for basal and apical channels at 50 pps.

The problem with these three studies is their use of the numerical pitch estimation procedure. Firstly, as mentioned in §2.3.4, it is prone to subject bias, and results can differ widely across subjects, as exemplified by Zeng's (2002) results. But the more serious flaw is that by only allowing a single number to describe each stimulus, the experimental design assumes a priori that the stimuli can be ordered in a single-dimensional perceptual space. This is an unwise assumption, as the stimuli are indisputably specified by two independent physical dimensions (electrode and rate), and it ignores the experimental evidence that cochlear implant place and rate are perceived independently (§3.8.1).
3.9 Conclusion

This chapter described cochlear implants and the percepts generated by electrical stimulation using simple sequences of pulses. The experimental evidence favours the view that cochlear implant place pitch and temporal pitch form two independent perceptual dimensions. Debate over these percepts sometimes hinges on the definition of pitch. None of the results cited in this chapter involved melody perception, which is crucial to the definition of pitch adopted in this thesis. Tests of melody perception have often been done with recipients listening to sound through their processors. Therefore, the next chapter will describe the methods used in processing sound to provide a useful sense of hearing to recipients. The discussion on pitch and melody perception will be resumed in chapter 5.
4 Cochlear implant sound processing

4.1 Introduction

This chapter describes the signal processing implementation of the three most widespread cochlear implant strategies: CIS, SPEAK and ACE. These strategies have much in common, and can all be described by the overall signal processing path shown in Figure 4.1. The front end amplifies the microphone signal to a suitable level. The filterbank splits the sound into multiple frequency bands, emulating the behaviour of the cochlea in a normal ear, where different locations along the length of the cochlea are sensitive to different frequencies. Each frequency band is allocated to one stimulation channel. The sampling and selection block samples the filter envelopes and determines the timing and pattern of the stimulation on each channel. The amplitude mapping block calculates the current level of each pulse according to the recipient's electrical dynamic range.

The following sections describe the signal path in more detail.

4.2 Microphones and front end

There are two types of microphones that could be used in a cochlear implant system. An omni-directional microphone has a single entry port for sound, and is equally sensitive to sound coming from all directions. It has a relatively flat frequency response. A directional microphone has two sound entry ports, and is most sensitive to sound coming from the front. Pre-emphasis is
an inherent feature of directional microphones. Nucleus cochlear implant systems use directional microphones.

The front end also implements automatic gain control (AGC) functions.

### 4.3 Filterbank

#### 4.3.1 Frequency allocation

The frequency bands are based on the function relating characteristic frequency to place in the normal ear. The default frequency allocation for the 22-channel filterbank on the Freedom processor is shown in Figure 4.2. The filter centre frequencies are linearly spaced below 1000 Hz, and logarithmically spaced above 1000 Hz.

![Figure 4.2 Freedom processor default 22-channel filterbank frequency response](image)

The following sections describe various filterbank implementations. All modern processors are implemented with digital signal processing (DSP), but the older switched capacitor filter (SCF) technology (a form of sampled analog signal processing) is mentioned briefly for completeness. The FIR filterbank is described in detail for later comparison to the less familiar FFT filterbank (§4.5).

#### 4.3.2 Switched-capacitor filterbank

The early prototypes of the SPEAK strategy for the Nucleus 22 cochlear implant system used a commercial SCF integrated circuit containing 16 sixth-order band-pass filters (McDermott et al. 1992). The Nucleus Spectra processor, introduced in 1994, had a custom integrated circuit
containing 20 fourth-order band-pass SCFs (Seligman and McDermott 1995). The Nucleus ESPrit behind-the-ear (BTE) processor, introduced in 1997, had the same SCF architecture as the Spectra, but with the entire processor implemented on one integrated circuit.

### 4.3.3 FIR filterbank

In a finite impulse response (FIR) digital filter, the output $y(n)$ is a weighted sum of the present and past values of the input signal $x(n)$:

$$y(n) = \sum_{m=0}^{M-1} h(m) x(n - m)$$  \hspace{1cm} (4.1)

The impulse response of the filter is $h(m)$. Each output sample $y(n)$ in Equation 4.1 requires $M$ multiply-accumulate operations. A structure to implement the filter is shown in Figure 4.3. A delay element is denoted by "$\Delta"$ and the summation by "$\Sigma$". The delay elements form a first-in, first-out (FIFO) data structure. In a software implementation, a circular buffer of $M$ memory locations is used.

![Figure 4.3 FIR filter structure, $M = 8$](image)

The frequency response of the filter is given by the Discrete Time Fourier Transform (DTFT):

$$H(\theta) = \sum_{m=0}^{M-1} h(m) e^{-j\theta m}$$  \hspace{1cm} (4.2)

The angular frequency variable $\theta$ is related to frequency in Hertz by the ADC sampling frequency $f_s$:

$$\theta = \frac{2\pi f}{f_s}$$  \hspace{1cm} (4.3)

For cochlear implant sound processing, a bank of band-pass filters is required. One method for designing a band-pass filter is to start with a low-pass filter $h(m)$, and modulate the coefficients,
i.e. multiply them by a sinusoid at the desired centre frequency of the filter. It is mathematically convenient to express this using a complex exponential, so that the filter with centre frequency $f_c$ has the impulse response:

$$g(m) = h(m) e^{\frac{j2\pi mf_c}{f_s}}$$  \hspace{1cm} (4.4)

The frequency response is:

$$G_i(\theta) = \sum_{m=0}^{M-1} g(m) e^{-j\theta m}$$

$$= \sum_{m=0}^{M-1} h(m) e^{\frac{j2\pi mf_c}{f_s}} e^{-j\theta m}$$  \hspace{1cm} (4.5)

$$= H\left(\theta - \frac{2\pi f_c}{f_s}\right)$$

i.e. $G(\theta)$ is a frequency-shifted version of $H(\theta)$, as shown in Figure 4.4. A complex filter can be implemented as a pair of real filters, where the in-phase filter uses the real part of the coefficients, and the quadrature filter uses the imaginary part:

$$\text{real}(g(m)) = h(m) \cos \left(\frac{2\pi mf_c}{f_s}\right)$$  \hspace{1cm} (4.6)

$$\text{imag}(g(m)) = h(m) \sin \left(\frac{2\pi mf_c}{f_s}\right)$$  \hspace{1cm} (4.7)

The frequency response of the quadrature filter has the same magnitude as the in-phase filter, but its phase lags by 90 degrees. Despite the apparent duplication in computation, a reason for
implementing both filters will be given in §4.4.3.

### 4.3.4 IIR filterbank

In an infinite impulse response (IIR) digital filter, the output also depends on past values of the output signal:

\[
y(n) = \sum_{m=0}^{M-1} b(m) \ x(n-m) - \sum_{m=1}^{M-1} a(m) \ y(n-m)
\]  

(4.8)

The original CIS implementation (Wilson et al. 1991b) used a bank of sixth-order IIR filters. IIR filters are usually considered more computationally efficient, because they can produce steep frequency response roll-offs with far fewer coefficients than FIR filters. However, because the present output depends on past values of the output (i.e. IIR filters are recursive), the output must be calculated at the same sampling rate as the input signal. In contrast, FIR filters are non-recursive; each output sample depends only on the past values of the input, so it is possible to calculate the output samples at a lower rate than the input sampling rate. The benefit of this will be shown in §4.4.3. IIR filters also carry the risks of instability and limit cycles, and require higher precision arithmetic.

### 4.4 Envelope detection

Representative signal waveforms of various envelope detection schemes are shown in Figure 4.5. The audio input is a short-duration tone burst (six cycles of a 375 Hz sinusoid). Figure 4.5b shows the response of a band-pass filter with centre frequency 375 Hz. The remaining plots show the outputs of the envelope detectors. Corresponding block diagrams are shown in Figure 4.6. Each diagram shows the processing for one channel.
Figure 4.5 Envelope detection signals.
(a) Band-pass filter output; (b) Peak follower output;
(c) Output of half-wave rectifier with 200 Hz smoothing filter;
(d) Output of full-wave rectifier with 400 Hz smoothing filter;
(e) Output of full-wave rectifier with 200 Hz smoothing filter;
(f) Quadrature envelope.
4.4.1 Peak follower

The Nucleus Spectra and ESPrit family of processors used peak followers (Seligman and McDermott 1995). In an equivalent DSP implementation, the output envelope $v(n)$ is calculated from the input signal $x(n)$ by:

$$
\begin{align*}
\text{if } x(n) & \geq v(n - 1) \\
v(n) &= x(n) \\
\text{else} & \\
v(n) &= a \cdot v(n - 1)
\end{align*}
$$

Thus if the input exceeds the output, the output follows the input, otherwise the output decays exponentially, with a time constant controlled by the coefficient $a$. The response to a pure tone has a ripple at the tone frequency (Figure 4.5b). The ripple could be reduced by increasing the decay time-constant, at the expense of increasing the rise and fall times.

4.4.2 Rectification and smoothing

Implementations of the CIS strategy typically use a Full Wave Rectifier (FWR) followed by a low-pass smoothing filter (Wilson et al. 1991a), typically with a cut-off frequency of 200 or 400 Hz (Loizou 1998). FWR with a 400 Hz cut-off produces ripple at twice the tone frequency (Figure 4.5d). The ripple can be reduced by lowering the filter cut-off to 200 Hz (Figure 4.5e).
The CIS strategy for the Clarion implant allowed a choice of FWR or HWR. With equal cut-off frequencies, the ripple is larger in HWR than FWR, and it modulates at the tone frequency (Figure 4.5c).

### 4.4.3 Quadrature envelope detection

Quadrature envelope detection for cochlear implant sound processing was first advocated by Eddington (1992), and was introduced commercially in the Nucleus SPrint processor in 1997. It is also used in the MED-EL TEMPO+ and OPUS processors (Zierhofer 1998; Helms et al. 2001), although they refer to it as the Hilbert transform method. It has been reported to improve speech perception compared to FWR and smoothing (Helms et al. 2001). As described in §4.3.3, each channel has a complex band-pass filter; i.e. a pair of band-pass filters having identical amplitude responses, and phase responses that differ by 90°. If we denote the real part of the complex filter output (i.e. the in-phase filter output) as \( x(n) \), and the imaginary part (i.e. the quadrature filter output) as \( y(n) \), then the envelope signal \( v(n) \) is simply the magnitude (Figure 4.6d):

\[
v(n) = \sqrt{x^2(n) + y^2(n)}
\]  

(4.10)

For a pure tone, the quadrature envelope has very little ripple (Figure 4.5f). This method is computationally efficient, because although a pair of filters is needed for each channel, the filter outputs only need to be calculated at the stimulation rate, rather than at the audio sampling rate. For example, Figure 4.7 shows the same 375 Hz tone as in Figure 4.5f. The in-phase (red) and quadrature (green) filter outputs, calculated at 16 kHz, are shown as small dots joined by continuous lines. If the stimulation rate is 500 Hz, then it is only necessary to calculate every 32nd output sample of each filter (shown as large dots). Even though a 500 Hz sampling rate is too low to represent either filter output signal individually, the quadrature envelope (large black dots) can be calculated at a 500 Hz rate because it is band-limited to 250 Hz (the filter bandwidth in this example).
This section describes the FFT filterbank implemented on the Nucleus SPrint and Freedom processors. The Discrete Fourier Transform (DFT) of an input sequence of \( M \) samples is:

\[
X(k) = \sum_{m=0}^{M-1} x(m) e^{-j2\pi mk/M} \quad \text{for } k = 0, \ldots, M-1
\]  

(4.11)

In contrast to a naïve implementation of the DFT, which requires \( M^2 \) multiply-accumulate operations, the Fast Fourier Transform (FFT) is a computationally efficient algorithm where the number of multiply-accumulate operations is \( 2M \log_2(M) \).
The processing for an FFT filterbank operates on a sliding buffer or FIFO of \( M \) samples, as shown in Figure 4.8 (note the similarity to Figure 4.3). The first step is to multiply the input samples by a window function \( h(m) \). The SPrint and Freedom use the Hann window, defined as:

\[
h(m) = 0.5 \left( 1.0 - \cos\left( \frac{2\pi m}{M} \right) \right) \quad \text{for } k = 0, \ldots, M - 1 \quad (4.12)
\]

The output is stored in a second memory buffer. The data in buffer location \( m \) at time index \( n \) is \( b(m,n) \), where:

\[
b(m,n) = h(m) \cdot x(n - m) \quad (4.13)
\]

Then an \( M \)-point FFT is performed on this buffer of data. The output of the FFT performed at time \( n \) is:

\[
B(k,n) = \sum_{m=0}^{M-1} b(m,n) \cdot e^{-j\frac{2\pi mk}{M}} = \sum_{m=0}^{M-1} h(m) \cdot x(n - m) \cdot e^{-j\frac{2\pi mk}{M}} \quad (4.14)
\]

Note that the transform sums over the index \( m \), and the output is stored in a buffer indexed by the frequency index \( k \). For \( k = 0 \), Equation 4.14 reduces to Equation 4.1, i.e. the first position in the output buffer contains an output sample of an FIR filter, which has an impulse response equal to the window function \( h(m) \). If we define:

\[
g_k(m) = h(m) \cdot e^{-j\frac{2\pi mk}{M}} \quad (4.15)
\]

then we can re-write Equation 4.14 as:

\[
u_k(n) = B(k,n) = \sum_{m=0}^{M-1} g_k(m) \cdot x(n - m) \quad (4.16)
\]

which represents the output of a bank of FIR filters (Harris 1982). There are \( M \) filters, and each FFT generates one new output sample of each filter. The \( k \)th filter has the complex impulse response given by Equation 4.15, which can be recognised as a band-pass filter in the same form as Equation 4.4, with centre frequency given by:

\[
f_c = \frac{k \cdot f_s}{M} \quad (4.17)
\]

From Equation 4.5, the frequency response of the \( k \)th filter is:
\[
G_k(\theta) = H\left(\theta + \frac{2\pi k}{M}\right)
\]

(4.18)

\(H(\theta)\) is a low-pass filter that has an impulse response equal to the window function \(h(m)\), and \(G_k(\theta)\) is a frequency-shifted version of \(H(\theta)\). As before, the complex filter can be expressed as a pair of quadrature filters:

\[
\text{real}(g_k(m)) = h(m) \cos\left(\frac{2\pi mk}{M}\right)
\]

(4.19)

\[
\text{imag}(g_k(m)) = -h(m) \sin\left(\frac{2\pi mk}{M}\right)
\]

(4.20)

The real part of each FFT bin represents one new sample of an \(M\)-point FIR filter, with centre frequency equal to the bin frequency. The imaginary part of each bin represents one new sample from the corresponding quadrature filter.

In the SPrint and Freedom, \(M = 128\), and the input sample rate is (approximately) 16 kHz. Bin 0 \((f_c = 0 \text{ Hz})\) and bin 64 \((f_c = 8000 \text{ Hz})\) are purely real, and bins 1 to 63 are complex. Because the input signal is real, the output has Hermitian symmetry, and bins 65 to 127 are not required. Thus there are 65 band-pass filters, with centre frequencies spaced linearly at multiples of 125 Hz. For the Hann window, the -6 dB bandwidth is 250 Hz (2 bins) (Harris 1978).

The FFT creates a bank of FIR filters with a linear frequency spacing of 125 Hz. However, a linear-log frequency spacing is required (§4.3.1). For the bands below 1000 Hz, each band can be simply assigned to one FFT bin, starting at bin 2 (centred at 250 Hz). Bins 0 and 1 are discarded. For the higher frequency bands, two or more consecutive FFT bins are combined to produce successively wider bands. There are two methods of combining bins: vector-sum and power-sum.

### 4.5.1 Vector-sum combination of bins

Regarding notation, the following expressions are independent of the time index \(n\), so it will be omitted; thus the complex output of the \(k\)th FFT bin will be denoted \(u(k)\), with real part \(x(k)\), and imaginary part \(y(k)\). The output channels will be indexed by \(m\). The \(m\)th composite channel is formed by a weighted vector sum of the FFT bins, giving a complex output \(s(m)\):

\[
s(m) = \sum_k \alpha(m, k) u(k)
\]

\[
= \sum_k \alpha(m, k) x(k) + j \sum_k \alpha(m, k) y(k)
\]

(4.21)

where \(\alpha(k, m)\) is a matrix of weights, which determines the frequency boundaries of the bands. Note that the weighted sum is carried out on the real and imaginary parts separately. Because FIR filtering is a linear operation, a weighted sum of FIR filter output samples is identical to the
output of a filter having an impulse response equal to the weighted sum of the individual impulse responses. In other words, the FFT vector-sum filterbank is truly an FIR filterbank; it produces identical outputs to the more traditional tapped delay line implementation (Figure 4.3). Thus the real part of \( s(m) \) is the in-phase output and the imaginary part is the quadrature output of a pair of FIR band-pass filters. Finally, the quadrature envelope is given by the magnitude:

\[
\nu(m) = \sqrt{s^2(m)} = \sqrt{\left( \sum_k \alpha(m,k) x(k) \right)^2 + \left( \sum_k \alpha(m,k) y(k) \right)^2}
\]  

(4.22)

### 4.5.2 Power-sum combination of bins

In the power-sum method, firstly the magnitude-squared (power) of each individual FFT bin is calculated. Then the power of each composite filter band is calculated as a weighted sum of the FFT bin powers, and finally the square-root is taken to give the envelope:

\[
\nu(m) = \sqrt{\sum_k a(k,m) \left( x^2(k) + y^2(k) \right)}
\]  

(4.23)

where \( a(k, m) \) is a matrix of weights, which again determines the frequency boundaries of the bands. In contrast to the vector-sum method, the power-sum envelope no longer corresponds exactly to the envelope of an FIR filter.

### 4.5.3 Advantages

The FFT filterbank implementation has three advantages over a direct form quadrature FIR implementation. The first is a saving in execution time. The calculation of one output of a single \( M \)-point real FIR filter requires \( M \) multiply-accumulate operations, thus a bank of \( 2N \) real filters (\( N \) quadrature pairs) requires \( 2MN \) operations. In comparison, the FFT vector-sum filterbank requires \( M \) multiplications for the windowing, \( 2M \log_2(M) \) operations for the FFT, and less than \( 2M \) multiply-accumulate operations for vector summation (there are \( M/2 \) complex bin values, i.e. \( M \) real values, each contributing to at most two bands). The total number of operations is approximately:

\[
N_{FFT} = M + 2M \log_2(M) + 2M = M(3 + 2 \log_2(M))
\]  

(4.24)

For \( N = 22 \) and \( M = 128 \), the FFT filterbank requires fewer multiply-accumulate operations than the FIR direct form implementation. The second advantage is that a lower number of multiply-accumulate operations reduces the amount of round-off noise due to fixed-point arithmetic. The third advantage is a substantial reduction in the memory required for the filter coefficients. The
direct form FIR requires $MN$ coefficients (assuming even-symmetry impulse responses). The FFT vector sum filterbank requires $M/2$ window coefficients, $M/4$ FFT "twiddle factors", and less than $2M$ weighting coefficients, for a total of less than $3M$. However, this does come at the expense of increased code space.

### 4.6 Responses to complex tones

This section describes the response of a filterbank to complex tones. The examples all use the 22-channel FFT vector-sum filterbank with the default ACE frequency boundaries, with the outputs calculated at 16000 Hz to show the temporal response in detail. The principles apply to any filterbank implementation.

Firstly consider a complex tone with two components, ramped on and off smoothly. If the two components are far apart in frequency, each will excite different filters, and the envelope of each channel will match the envelope of the input tone. Figure 4.9 shows the lowest eight channels of the 22-channel filterbank in response to components at 500 and 1000 Hz (which lie at the centres of the third and seventh filters). The filter centre frequencies are indicated on the left hand axis. Two groups of channels are activated, separated by one inactive channel; thus the two components are resolved by the filterbank.

![Figure 4.9 Filterbank response to two resolved components.](image)

Red: filter output; Black: quadrature envelope

In contrast, two components will not be resolved if they both lie within the pass band of one filter. Figure 4.10 shows the response for components at 500 and 625 Hz. A single group of four adjacent channels are activated, so it is no longer clear that there are two components. The third filter (and also the fourth) responds to both components, so its output is no longer sinusoidal.
The amplitude of the envelope modulates at the difference frequency of 125 Hz. Note that spectral resolution is determined by the bandwidth of the filters (the -6 dB bandwidth is 250 Hz) rather than the spacing of the centre frequencies: if additional filters (having the same bandwidth) were placed midway between each of the existing filters, the resolution would be unchanged.

Finally consider a harmonic tone with fundamental frequency, F0. The difference in frequency between successive harmonics is F0, so if more than one harmonic falls within the bandwidth of a filter, the envelope modulation frequency will be F0. This provides a potential temporal pitch cue. Figure 4.11 shows the filterbank response to a sequence of four harmonic tones having F0s equal to the centre frequencies of the first four filters. The top panel displays the envelopes as a 22-channel spectrogram. The bottom panel plots the envelope waveforms of the first nine channels. In both panels, the filter centre frequencies are indicated on the left hand axis, and the F0 of the tone is shown above each stimuli. Each tone has the same spectral profile (this profile was used in the Modified Melodies test that will be reported in chapters 11 and 12). For the 125 Hz tone, each channel responds to multiple harmonics, and has an envelope that modulates at 125 Hz; the harmonics are unresolved. For the 250 Hz tone, the harmonics exactly coincide with the centre frequencies of the odd-numbered filters; those filters respond to a single harmonic and their outputs do not modulate. The even-numbered filters respond to two harmonics and thus modulate at 250 Hz. Thus the harmonics are partially resolved. For the 375 Hz tone, the harmonic spacing is greater than the filter bandwidth, so every filter responds to a single harmonic, and there is no envelope modulation. The harmonics are resolved: channels two and five (centred on the first two harmonics) are larger in amplitude than their neighbours, as is
evident in the spectrogram. For the 500 Hz tone, the first two harmonics are very clearly resolved, with two separate groups of activated channels.

The last example is Figure 4.12, which shows the response to a sequence of four harmonic tones that are only one semitone apart, starting at C4 (262 Hz). In this frequency range the harmonics are partially resolved, but some envelope modulation occurs on the channels that lie between the harmonics. In examining the spectrogram, the direction of the F0 change is not immediately apparent. Small increases in F0 produce amplitude changes that move in opposite directions on neighbouring channels. The waveform plot shows that channel 3 (centred at 500 Hz) decreases in amplitude as F0 increases, whilst channel 4 (centred at 625 Hz) increases in amplitude. It appears that place cues may not be a useful guide to F0 for such harmonic tones.

It should also be noted that the audio waveforms of the example tones in Figure 4.11 and Figure 4.12 were normalised to have the same amplitude, and they all have about the same loudness to a normal hearing listener. However, the resulting channel amplitudes vary considerably across the eight stimuli, as evident in the spectrograms. The signal power is spread evenly across channels for the 125 Hz tone, giving low channel amplitudes; in contrast the power is concentrated in two peaks for the 500 Hz tone, giving higher channel amplitudes. It is difficult to predict the loudness relationships between the resulting stimuli because of the subsequent amplitude compression (§4.8), the relationship of current to loudness (§3.5.1), and the summation of loudness across channels (§3.5.2). However, it is possible that a small change in F0 could produce a significant loudness change.
Figure 4.11 Filterbank response to harmonic tones.
Top: Spectrogram. Bottom: Envelope waveforms for nine channels.
Figure 4.12 Filterbank response with small F0 changes.
Top: Spectrogram. Bottom: Envelope waveforms for four channels.

4.7 Sampling and selection

The filterbank provides a set of parallel envelope signals. In contrast, the implant stimulates sequentially. It is therefore necessary to sample the envelopes and select one channel at a time
for a pulse. This sampling and selection process is where the CIS, SPEAK, and ACE strategies differ most. CIS is the simplest of these three strategies, so it will be described first.

### 4.7.1 CIS

The Continuous Interleaved Sampling (CIS) strategy (Wilson et al. 1991a) was developed for the Ineraid implant, which had six intracochlear electrodes, directly connected to a percutaneous plug. The filterbank thus had six frequency bands. The six filter envelopes were each sampled at a fixed rate, with the sample times (and thus the pulses) on each channel interleaved in a round-robin fashion. This yields the pulse sequence that was shown in Figure 3.3 (and Figure 3.4). The stimulation rate on each channel was high enough to avoid any rate-related pitch cues, typically over 800 pps.

### 4.7.2 SPEAK and ACE

The Spectral Maxima Sound Processor (SMSP) strategy was developed at the University of Melbourne for the Nucleus 22 implant (McDermott et al. 1992). It used a 16-channel filterbank (§4.3.2). This was increased to 20 channels in the Spectra processor, and the strategy became known as SPEAK.

The total stimulation rate achievable with the Nucleus 22 implant is dependent on the recipient's current levels, but is usually less than 2000 pps. Typically BP+1 mode is used, which allows 20 channels (§3.4). If each envelope was sampled and stimulated in turn (as in CIS), the rate on each channel would be less than 100 pps, which is too slow to convey the amplitude changes adequately. Instead, the SPEAK strategy operates at a 250 Hz analysis rate. In each analysis period (4 ms), it samples all 20 filter envelopes, finds the six envelopes having the largest amplitude (“maxima”), and sequentially presents one pulse on each of the six corresponding electrodes. Thus the electrodes stimulated in each analysis period vary as the sound spectrum changes, i.e. the stimulation “roves” across the electrode array. An electrode that is repeatedly selected in consecutive analysis periods is stimulated at 250 pps (e.g. during the steady portion of a vowel).

Higher rates became available with the introduction of the Nucleus 24 implant. The ACE strategy uses the same maxima selection algorithm as SPEAK, but the clinical software allows a wider choice of stimulation rates and number of maxima.

### 4.8 Amplitude mapping

The electrical dynamic range between T and C level is typically only 10 dB. Therefore the much larger dynamic range of the filterbank envelope signals must be compressed into the recipient’s
electrical dynamic range. In the Nucleus system, this compression is described by the loudness growth function (LGF; Figure 4.13):

\[
y = \begin{cases} 
\log \left( \frac{1 + \alpha \left( \frac{v - B}{M - B} \right)}{\log(1 + \alpha)} \right) & B \leq v \leq M \\
0 & v \leq B \\
1 & v \geq M 
\end{cases}
\]

(4.25)

where \( v \) is the envelope, \( B \) is the base level, \( M \) is the saturation-level, and the parameter \( \alpha \) controls the steepness of the compression curve. The output \( y \) is the proportion of the electrical dynamic range, expressed in clinical current level units.

![Figure 4.13 Non-linear compression function](image_url)

The compressed amplitude \( y \) is then linearly scaled to give a current level \( c \) (in clinical units) according to the equation:
\[ c = T + (C - T) \gamma \]  

(4.26)

where \( T \) is the threshold level and \( C \) is the maximum comfortable level of the corresponding channel, expressed in clinical units. Thus an envelope equal to the base-level \( B \) (Figure 4.13) produces a current at the \( T \) level. Envelope values smaller than this are discarded. Envelope values equal to or greater than the saturation level \( M \) result in stimulation at the \( C \) level.

In acoustic hearing, loudness is a power function of sound pressure \( p \) with exponent \( \beta_a = 0.6 \) (Equation 2.6, §2.3.4) (Stevens 1986). Fu and Shannon (1998) found that in electrical stimulation, loudness was a power function of current \( i \) with exponent \( \beta_e = 2.7 \) (Equation 3.3, §3.5.1). They reasoned that to match the loudness between acoustic and electric hearing:

\[ L = k_a p^{\beta_a} = k_e i^{\beta_e} \]  

(4.27)

requires current to be related to sound pressure (and hence microphone voltage and envelope amplitude) by another power law:

\[ i = k_m p^{\beta_m} \]  

(4.28)

where the exponent is \( \beta_m = \beta_a / \beta_e = 0.6 / 2.7 = 0.22 \). Fu and Shannon (1998) measured phoneme recognition with a four-channel CIS processor in three Nucleus 22 recipients, using a power-law amplitude mapping function with exponent \( \beta \) ranging from 0.05 to 0.75, and found best recognition for \( \beta = 0.2 \). Performance was fairly insensitive to the value of the exponent in the range 0.1 to 0.5, and they observed that the Nucleus logarithmic-shaped compression function was a close approximation to a power-law with exponent 0.25.

It should be noted that no commercial sound processing strategy accounts for the effects of loudness summation across channels (§3.5.2). The SpeL research strategy was developed to address this issue (McDermott et al. 2005).
5 Cochlear implant pitch perception

5.1 Introduction

Chapter 3 introduced the two aspects of cochlear implant pitch perception: place and temporal cues. This chapter firstly examines the procedures that have been used to measure pitch and melody perception. It then considers the evidence on whether cochlear implant place pitch is capable of conveying melodies.

5.2 Testing pitch and melody perception

This section reviews the procedures that have been used by other researchers to investigate pitch and melody perception with cochlear implant recipients. For reference, Table 5.1 lists the test procedures that were used in some selected studies. Some studies with acoustic hearing are also included for comparison.
### Table 5.1 Pitch and melody perception tests used in some selected studies

<table>
<thead>
<tr>
<th>Test procedure</th>
<th>Acoustic Hearing</th>
<th>Cochlear Implants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tone discrimination</td>
<td>(Moore and Peters 1992)</td>
<td>(Gfeller et al. 2002)</td>
</tr>
<tr>
<td></td>
<td>(Hyde and Peretz 2004)</td>
<td>(Gantz et al. 2004)</td>
</tr>
<tr>
<td>Tone pitch ranking</td>
<td></td>
<td>(Geurts and Wouters 2001)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Gfeller et al. 2002)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Gantz et al. 2004)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Laneau et al. 2004)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Laneau et al. 2006)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Nimmons et al. 2008)</td>
</tr>
<tr>
<td>Multidimensional scaling</td>
<td>(Plomp 1976, p108)</td>
<td>(Tong et al. 1983a)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Tong et al. 1983b)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(McKay et al. 1996)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Collins and Throckmorton 2000)</td>
</tr>
<tr>
<td>Pitch scaling</td>
<td></td>
<td>(Eddington et al. 1978)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Busby et al. 1994)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Cohen et al. 1996)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Busby and Clark 1997)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Fearn 2001, chapter 3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Zeng 2002)</td>
</tr>
<tr>
<td>Melody discrimination</td>
<td>(Trainor 1996)</td>
<td>(Büchler et al. 2005)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Vongpaisal et al. 2006, experiment 4)</td>
</tr>
<tr>
<td>PMMA tonal subtest</td>
<td>(Gordon 1979)</td>
<td>(Gfeller and Lansing 1991)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Gfeller and Lansing 1992)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Gfeller et al. 1997)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Fearn 2001, chapter 8)</td>
</tr>
<tr>
<td>Montreal Battery for Evaluation of Amusia</td>
<td></td>
<td>(Peretz et al. 2003)</td>
</tr>
<tr>
<td>Open-set melody identification</td>
<td>(Burns and Viemeister 1981)</td>
<td>(Eddington et al. 1978)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Pijl and Schwarz 1995)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Gfeller et al. 2002)</td>
</tr>
<tr>
<td>Closed-set melody identification</td>
<td>(Dowling and Fujitani 1971)</td>
<td>(Pijl and Schwarz 1995)</td>
</tr>
<tr>
<td></td>
<td>(Burns and Viemeister 1976)</td>
<td>(Stordahl 2002)</td>
</tr>
<tr>
<td></td>
<td>(Moore and Rosen 1979)</td>
<td>(Kong et al. 2004)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Kong et al. 2005)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Olszewski et al. 2005)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Chen and Zeng 2005)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Laneau et al. 2006)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Spahr et al. 2007)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Mitani et al. 2007)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Nimmons et al. 2008)</td>
</tr>
<tr>
<td>Musical interval estimation</td>
<td></td>
<td>(McDermott and McKay 1997)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Laneau et al. 2006)</td>
</tr>
<tr>
<td>Musical interval judgements</td>
<td>(Burns and Viemeister 1981)</td>
<td>(Pijl 1997a)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(McDermott and McKay 1997)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Fearn 2001, chapter 5)</td>
</tr>
<tr>
<td>Distorted Tunes Test</td>
<td>(Fry 1948)</td>
<td>(Galvin et al. 2007)</td>
</tr>
<tr>
<td></td>
<td>(Kalmus and Fry 1980)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(Drayna et al. 2001)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(Ayotte et al. 2002)</td>
<td></td>
</tr>
</tbody>
</table>

### 5.2.1 Tone discrimination and ranking

The limitations of discrimination tasks (§2.3.2) are particularly acute for cochlear implant pitch studies. A cochlear implant recipient asked to discriminate between two tones with differing
fundamental frequencies may detect differences that are not pitch at all. As was explained in §4.6, even if the sound waveforms have the same amplitude, a change in fundamental frequency can change the resulting amplitudes of the stimulation pulses in a manner that affects loudness. Thus, in a NAFC task, the variant stimulus may have an unintentional loudness difference that helps the subject to detect it. Furthermore, in a pitch discrimination task, the variant tone may have a higher fundamental frequency than the reference, yet the subject may perceive it as being lower in pitch. This pitch reversal would go undetected in a discrimination task, but would be immediately apparent in a ranking task. Such pitch reversals have been observed in cochlear implant pitch ranking studies (Vandali et al. 2005; Sucher and McDermott 2007). It is clearly unwise to infer good pitch perception from a high tone discrimination score.

A pitch reversal means that perceived pitch is a non-monotonic function of fundamental frequency. An essential prerequisite for the use of an adaptive procedure (§2.3.3) is that the underlying psychometric function must be monotonic. If it is not, an adaptive procedure may fail to converge, or may converge to an erroneous value. This crucial point has been neglected in several cochlear implant pitch perception studies (e.g. Geurts and Wouters 2001; Fearn 2001; Gfeller et al. 2002; Mitterbacher et al. 2005; Nimmons et al. 2008). Cochlear implant pitch-ranking experiments should use the method of constant stimuli (§2.3.3), which directly measures the shape of the psychometric function. The virtue of this method will be demonstrated in the experiments in chapters 8, 9, and 10 of this thesis.

In a study of 12 cochlear implant recipients, Gfeller et al. (2002) found no correlation between the results of pitch ranking with complex tones, and 4AFC discrimination with pure tones. This was most likely due to the use of different cues for pure and complex tones, but any comparison is confounded by the different procedures used: subjects who scored better on pure tone discrimination may have used non-pitch cues or had abnormal pitch ordering. Furthermore, both tasks used adaptive procedures, and results may have been affected by convergence problems.

Although ranking tasks are preferable to discrimination tasks, there is still no guarantee that subjects are actually using pitch in their judgments. Subjects may base their decision on some other aspect of the sound that can be ranked. It is plausible that this could be brightness, which also can be ordered on a low-to-high scale (§2.5) (Lichte 1941). Many normal hearing listeners are willing to describe a change in spectral profile as a pitch change (§2.5.1) (Pitt and Crowder 1992).

**5.2.2 Melody discrimination**

The "Primary Measures of Music Audiation" (PMMA) is a widely used test of musical aptitude, initially developed for use with young children (Gordon 1979). The tonal subtest measures the ability to discriminate pitch patterns. Each stimulus consists of a short sequence of up to five
notes played on a piano. In each trial, a pair of sequences is presented, and the subject is asked whether the two sequences were the same or different. In half of the trials, the two sequences match, and in the remaining trials the two sequences have identical rhythm, but differ in pitch. The PMMA was used in several studies of cochlear implant recipients (Gfeller and Lansing 1991; Gfeller and Lansing 1992; Gfeller et al. 1997; Fearn 2001). Scores of cochlear implant recipients were significantly lower than normal hearing listeners. Fearn (2001, chapter 8) saw no difference in performance for pure tones and the usual piano tones. He noted that the PMMA tonal subtest was a test of pitch discrimination, and not melody perception per se. In his results, 74% of the cochlear implant recipient errors were a "same" response when the sequences were different, which strongly suggested that recipients had insufficient pitch resolution.

The Montreal Battery for Evaluation of Amusia (MBEA) was developed to diagnose disorders in musical ability acquired congenitally or as a result of brain injury (Peretz et al. 2003). It comprises six subtests, three of which are concerned with pitch. It uses a same/different melody discrimination task, similar to the PMMA tonal subtest.

Trainor (1996) used a yes/no vigilance task. The standard melody consisted of the five-note sequence C4-E4-G4-E4-C4. A background sound was created by continuously repeating this melody, with pseudo-random transpositions. Every now and then a trial occurred, in which the third note of the melody was lowered in pitch by a quarter of a semitone. The subject’s task was to respond when a different melody was heard. This procedure was originally developed for use with infants, who would respond by turning their head towards the loudspeaker. Trainor (1996) also used it with normal hearing adults to measure the effect of harmonic content on pitch discrimination. The main drawback of this procedure is its inefficiency, due to the large number of repeated presentations of the standard melody. The number of standard presentations between each experimental trial must vary pseudo-randomly, and was at least two, so the total number of presentations was more than three times the number of trials. Furthermore, the score in a yes/no test depends on the strictness of the subject’s criterion for responding (i.e. the results are affected by subject bias, §2.3.1). This criterion will vary from subject to subject, and may vary from session to session in the same subject. To address this issue, Trainor estimated the sensitivity $d’$ from the hit rate and false alarm rate, but a forced-choice paradigm is preferable.

A cochlear implant music test battery developed at University Hospital Zurich (Büchler et al. 2005) included a same/different task with pairs of melodies differing in a single note. It was reported to be too difficult for most cochlear implant recipients, with most failing to score above chance. It was planned to omit this test from subsequent versions of the battery.

Melody discrimination tasks are unsatisfactory for cochlear implant recipients for several reasons. As mentioned in §2.3.2, there could be unintended non-pitch differences, or pitch reversals. Even in the absence of those anomalies, a discrimination task provides no evidence
that the melody is being perceived in the same way as a normal hearing listener perceives it. A reference melody and its variant may sound different to each other, but neither may be recognisable. For example, the Nucleus cochlear implant fitting software allows the number of electrodes allocated to the lower frequency range to be increased. With this expansion of the frequency scale, subjects may be better able to detect the difference between the reference and variant melodies, yet neither may have the desired musical interval relationships. A preferable test is one that provides evidence of correctly perceiving musical intervals.

### 5.2.3 Familiar melody identification

In an open-set melody identification test (sometimes called a melody recognition test), the subject hears a melody and must name it. Gfeller et al. (2002) used a set of 12 melodies selected to be familiar to American adults. The scores of cochlear implant recipients were much worse than those of normal hearing subjects, and they found the task very demanding.

To compare the performance of one subject under different conditions (e.g. with different sound processing strategies), a closed set should be used, i.e. the subject hears the melody then is asked to select the melody name from a list of possible answers. Closed-set melody identification is perhaps the most commonly-used test of cochlear implant pitch perception, as demonstrated by the number of studies listed in Table 5.1.

The first issue with melody identification (either open or closed set) is that rhythmic cues alone may be sufficient to identify the melodies. In Gfeller et al. (2002), the cochlear implant recipients obtained significantly better scores for those melodies containing distinctive rhythmic patterns. To avoid rhythm cues, Dowling and Fujitani (1971) used a set of five melodies that had identical rhythm; however this technique shrinks the pool of available melodies. A more common approach is to split the longer duration notes into a succession of shorter notes, creating *isochronous* melodies (e.g. Burns and Viemeister 1976; Moore and Rosen 1979; Kong et al. 2005; Nimmons et al. 2008).

The second issue is that it can be difficult to find an adequate set of melodies that are familiar to a cochlear implant recipient, whom may have very little musical experience. The need to remove rhythmic cues exacerbates the problem by making those melodies sound less familiar. There are several approaches to dealing with this issue. After completing the open-set test of Gfeller et al. (2002), each subject was shown a printed list of the names of the melodies, and asked whether they were familiar with them. Their melody recognition score was then adjusted to exclude the unfamiliar melodies. In a closed-set test, the subject can be asked about melody familiarity beforehand, and the unfamiliar melodies omitted from the set. Another issue is that lists of familiar melodies are culturally specific, and cannot generally be re-used across countries.
A less obvious issue is that some normal hearing listeners are able to obtain high identification scores in the face of deliberate pitch distortions. As discussed in §2.7.2, Dowling and Fujitani (1971) reported that subjects averaged 99% correct for undistorted melodies, but still scored significantly above chance for two conditions where the melodies were modified: 66% when both contours and the relative sizes of the musical intervals were preserved, and 59% when only the contours were preserved (chance was 20%). Moore and Rosen (1979) tested 20 normal hearing subjects on a set of ten isochronous melodies. In one condition, all the musical intervals were doubled, i.e. a one semitone step was replaced with a two semitone step, and so on. Despite this drastic pitch alteration, nine subjects obtained scores significantly better than chance, including two subjects who achieved scores of 90% or more. In another condition, all the musical intervals were compressed, with an octave (a frequency ratio of 2) reduced to a frequency ratio of 1.3 (about 4.5 semitones), and seven subjects scored significantly better than chance. This type of manipulation disturbs the musical intervals, but preserves the melodic contour and the relative sizes of the intervals (§2.7.2). These results show that melodic contour alone is sufficient for some normal hearing subjects to recognise a melody. It also demonstrates that a high score on a melody identification task does not necessarily imply accurate perception of musical intervals, and so melody identification may not be a sensitive test for comparing cochlear implant sound processing strategies.

5.2.4 Melodic Contour Identification

The Melodic Contour Identification (MCI) task was recently developed specifically for testing music perception in cochlear implant recipients (Galvin et al. 2007). Nine five-note sequences were constructed, as illustrated in Figure 5.1. Each note was a complex tone, consisting of the first three harmonics, with amplitudes 0, -3, and -6 dB. The musical interval between successive notes varied between one and five semitones (giving an overall range for the "rising" and "falling" contours of between four and twenty semitones), and the lowest note was one of A3 (220 Hz), A4 (440 Hz) or A5 (880 Hz). The set of stimuli thus contained 135 sequences (9 contours x 5 intervals x 3 base notes). The subject performed a closed-set identification task: in each trial, one sequence was presented, and the subject was asked to select one of nine on-screen representations of the contours.
Galvin et al. reported results for 11 cochlear implant recipients, who used a variety of implant types, processing strategies, and pulse rates. As might be expected, performance increased with increasing numbers of semitones between notes. Performance generally was slightly worse for the lowest base note (A3).

The main objection to this procedure is that it suffers from the same perceptual ambiguity as a pitch-ranking task: there is no guarantee that the recipients were using melodic pitch to identify the contours. Because each note had three harmonic components, strong place pitch cues were available. If cochlear implant place-pitch is perceived similarly to brightness in normal hearing, then a recipient could obtain a high score in the MCI task, even if the contours did not provide any sensation of melody. This is suggested by the finding that MCI scores were significantly correlated with closed-set vowel identification scores, but were not correlated with familiar melody identification scores.

### 5.2.5 Musical interval judgements

McDermott and McKay (1997) tested musical interval perception with a cochlear implant recipient who had trained as a piano tuner before becoming deaf. The recipient heard two notes repeatedly, and was asked to adjust the pitch of the second note to obtain a specified musical interval, e.g. a fifth (7 semitones). He performed this task with good accuracy when the notes were presented by varying the pulse rate on a single electrode.
For musically untrained subjects, Pijl (1997a) used familiar melodies as exemplars of musical intervals. For example, the first four notes of Twinkle Twinkle Little Star contain a rising fifth (CCGG). The subject was asked to mentally rehearse the chosen melody, and was then presented with a sequence of four notes (CCxx), where the last two notes were shifted up or down in pitch together. The subject was asked whether the final notes were too high, correct, or too low. The twelve notes in the octave above the starting note were presented in randomised order. Two cochlear implant recipients performed the task with the notes presented by varying the pulse rate on a single electrode. They obtained similar accuracy to normal hearing subjects listening to piano sounds. In contrast, the cochlear implant recipients performed at chance levels when listening to piano sounds processed by the SPEAK strategy.

A similar procedure was used by Fearn (2001, chapter 5). In each trial, the subject heard an initial phrase of a familiar melody, using synthesised piano notes, with the final note shifted up or down in pitch. The subject was asked whether the final note was too high, too low, or correct. An adaptive procedure was used to find the note that the subject judged to be correct. Two randomly interleaved staircases were used, with starting points 13 semitones above and below the correct note, and with a step size of four semitones initially, reducing to one semitone after three reversals. A group of 21 normal hearing listeners participated in the study, and most converged rapidly to the correct note. In contrast, a group of 25 cochlear implant recipients (listening to piano sounds through their usual sound processor) showed much poorer accuracy, with a mean final error of three semitones. One criticism of this test is that an adaptive procedure was used. As mentioned in §5.2.1, an adaptive procedure may fail to converge in the presence of pitch reversals, which have been observed in cochlear implant recipients with harmonic tones.

5.2.6 Distorted Tunes Test

The Distorted Tunes Test was developed in the 1940s (Fry 1948; Kalmus and Fry 1980) to detect people who were “tune deaf”. The test used 25 familiar melodies, with rhythm cues intact. For each melody, a distorted version was prepared by changing several notes by one or two semitones, preserving the overall contour of the melody. The stimulus set comprised the 25 correct and the 25 distorted melodies. In each trial, one melody from the set was presented, and the subject was asked whether the melody was correct or distorted. It is thus classed as a single interval (yes/no) task (§2.3.1), and so can be affected by subject bias. Given that cochlear implant pitch perception is generally very poor, it is quite possible that some cochlear implant recipients would respond “distorted” to every presentation, and the test would provide very little information.

In the original version of the test, each melody in the set was presented once (in randomised order), for a total of 50 trials. Subsequent analysis of the results showed that only 26 out of the
50 trials significantly contributed to the discrimination between normal and abnormal melody perception, so the remaining 24 trials were omitted. This final version of the Distorted Tunes Test had the drawback of not being balanced: there were 17 distorted presentations and only 9 correct presentations.

The melodies used by Fry and Kalmus were appropriate for the British population in the 1940s. Drayna (2001) adapted the test for Americans by selecting a new set of melodies and distortions (an on-line version of this test is available at the NIDCD web-site, http://www.nidcd.nih.gov/research/scientists/draynad.asp). Both versions included melodies that would be unfamiliar to many Australians.

5.3 Is place pitch really pitch?

In the electrode pitch-ranking study described in §3.6, Nelson et al. (1995) added this caveat:

The term "pitch" is used here with some caution. No one has demonstrated melody recognition for across-electrode stimulations, which should be a requirement for the percept to be truly called "pitch." … Some subjects report that stimuli sound "sharper" on adjacent electrodes, which may indicate that the appropriate perceptual dimension correlated with stimulation across electrodes is "timbre" rather than pitch. The procedures used here do not distinguish between the dimensions of "pitch" and "timbre."

More than a decade later, there remains scant evidence of melody perception by cochlear implant recipients using place cues alone.

As mentioned in §5.2.5, McDermott and McKay (1997) conducted a study with a musically-trained cochlear implant recipient. Two single channels of his implant were stimulated in succession (at a fixed pulse rate of 250 pps), and the subject was asked to name the musical interval between them. With a finite set of electrodes, having a fixed spacing of 0.75 mm, only a discrete set of pairings were possible, which were unlikely to exactly correspond to standard musical intervals. Nevertheless, the subject was able to perform the task consistently, naming larger intervals in response to stimuli having larger channel separations. Note that (unlike in a rate-pitch experiment) there is no objective way of determining the "correct" interval for any pair of stimuli. One criticism of the experiment is that the low pulse rate used (250 pps) would have evoked a rate-pitch percept that (despite being constant) may have been difficult to ignore. This was perhaps justifiable as the recipient was a user of the SPEAK strategy (which stimulated at 250 pps), but it would have been better to use a much higher pulse rate to minimise any concurrent rate pitch cues (the recipient's Nucleus 22 implant allowed at least 1500 pps on a single channel). In another condition, the two stimuli differed in both channel and pulse rate (in the range 100 to 200 pps). There was some interaction of the two cues, but the
place cue appeared to dominate.

Laneau (2005, chapter 7; Laneau et al. 2006) tested four cochlear implant recipients in a closed-set melody identification task, played on a synthesised MIDI clarinet, without rhythm cues. Melodies played in a low and a high frequency range were tested separately (having a median note of either 185 or 370 Hz). Three of the subjects scored significantly above chance with the ACE strategy for the high frequency ranges (scores were approximately 40 – 60%, chance was 10%). Harmonic tones in this frequency range produce little amplitude modulation (§4.6), and subjects are less sensitive to temporal cues above 300 Hz, which could imply that the subjects were relying on place pitch cues. Surprisingly, there was no statistically significant difference in scores between the low and high frequency ranges. It is possible that the higher frequency range melodies did provide some residual temporal cues, especially for the lower notes of the melodies, which ranged down to 262 Hz (C4) (see Figure 4.12 for examples of tones with F0s from 264 to 311 Hz).

In an unpublished study, Chen and Zeng (2005) investigated closed set melody identification by cochlear implant recipients listening to synthesised pulse trains. Twelve melodies were used, with rhythmic cues removed. In the temporal pitch condition, a single electrode was stimulated at a pulse rate equal to the fundamental frequency of the note. In the place pitch condition, all notes had the same pulse rate, and the stimulating electrode was varied. Ideally each note would be assigned to a different electrode, but the electrode spacing corresponds to too large a frequency difference. Instead, groups of successive notes (either 6 or 3 semitones) were all assigned to the same electrode. Scores in the place-pitch condition were comparable to scores in the temporal-pitch condition, and one subject obtained a score of 80% correct with 6 consecutive semitones assigned to each electrode. Clearly, in the place-pitch condition, the musical intervals were not correct, and the melodic contour was very coarsely represented. Although at first these results appear surprising, they are reasonably consistent with the previously described results (§2.7.2) of Dowling and Fujitani (1971), where normal hearing subjects averaged 59% on identification of five familiar melodies (without rhythm cues) when only the pitch contours were preserved. The Chen and Zeng (2005) study implies that place-pitch alone was sufficient to convey some melodic pitch contour information.

An experiment to directly address the question of whether cochlear implant place pitch can support melody perception will be described in Chapter 13.
6 Improving cochlear implant pitch perception

6.1 Introduction

This chapter provides a survey of schemes that have been proposed to improve pitch perception. Because of the commercial nature of this field, patents can be a good source of technical information. All the patents and patent applications referred to are listed in Table 6.1. The final column provides cross references. The date shown is the earliest priority date for the patent application. No opinion as to the validity or infringement of any patent is implied.

<table>
<thead>
<tr>
<th>Inventors</th>
<th>Number</th>
<th>Priority Date</th>
<th>Subject</th>
<th>Section</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wolfe, Joe Carter, Paul Parker, Simon Fearn, Robert Frampton, Niki</td>
<td>US 7072717 B1</td>
<td>1999-07-13</td>
<td>Multirate</td>
<td>6.3.3</td>
</tr>
<tr>
<td>Zierhofer, Clemens</td>
<td>US 6594525 B1</td>
<td>1999-08-26</td>
<td>Channel-Specific Sampling Sequences</td>
<td>6.3.6</td>
</tr>
<tr>
<td>Blamey, Peter Swanson, Brett McDermott, Hugh Patrick, Jim Clark, Graeme</td>
<td>US 7082332 B2</td>
<td>2000-06-19</td>
<td>Travelling Wave</td>
<td>6.2.5</td>
</tr>
<tr>
<td>Swanson, Brett McDermott, Hugh Blamey, Peter Patrick, Jim</td>
<td>WO 2004/021363 A1</td>
<td>2002-09-02</td>
<td>Half Wave Gating</td>
<td>6.2.2</td>
</tr>
<tr>
<td>McDermott, Hugh McKay, Colette</td>
<td>US 7231257 B2</td>
<td>2003-02-28</td>
<td>F0-Sync</td>
<td>6.4.2</td>
</tr>
</tbody>
</table>

Table 6.1 Patents and patent applications

The sound coding strategies can firstly be classified according to whether they seek to enhance place or temporal pitch cues. The large variety of temporal pitch strategies can be further subdivided into a $2 \times 2$ matrix of categories, as listed in Table 6.2. The first aspect of this classification scheme is whether the strategy conveys temporal cues by amplitude modulation of
a fixed high-rate carrier, or by controlling the pulse timing (i.e. pulse rate). As was reviewed in §3.7, the perceptual effects of varying modulation frequency and pulse rate are similar, so both approaches have merit. The second aspect of the classification scheme is whether the strategy provides a temporal cue within each channel independently, or whether it attempts to provide a "global" temporal cue across all of the channels together.

<table>
<thead>
<tr>
<th>Amplitude modulation cues</th>
<th>Pulse timing cues</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Within-channel</strong></td>
<td></td>
</tr>
<tr>
<td>Half wave rectification</td>
<td>Peak Picker</td>
</tr>
<tr>
<td>Half wave gating</td>
<td>Peak Derived Timing</td>
</tr>
<tr>
<td>F0-CIS</td>
<td>Multirate</td>
</tr>
<tr>
<td>Modulation Depth Enhancement</td>
<td>Spike-based Temporal Auditory Representation</td>
</tr>
<tr>
<td></td>
<td>Asynchronous Interleaved Sampling</td>
</tr>
<tr>
<td></td>
<td>Channel-Specific Sampling Sequences</td>
</tr>
<tr>
<td><strong>Global</strong></td>
<td></td>
</tr>
<tr>
<td>F0 Modulation</td>
<td>Nucleus 22 F0 strategies</td>
</tr>
<tr>
<td>Multichannel Envelope Modulation</td>
<td></td>
</tr>
</tbody>
</table>

Table 6.2 Temporal pitch strategies

### 6.2 Within-channel amplitude modulation cues

It was shown in §4.6 that unresolved harmonics produce amplitude modulation of the filter envelopes. This can provide an important cue to pitch in CIS and ACE. The strategies in this category attempt to enhance this amplitude modulation cue on each channel independently.

#### 6.2.1 Half wave rectification (HWR)

In the CIS, SPEAK and ACE strategies, the amplitude of the signal in each frequency band is estimated by an envelope detector (§4.4), eliminating the phase information (sometimes referred to as the temporal fine structure). This contrasts with normal hearing, where the inner hair cells preferentially respond to deflections of the basilar membrane in one direction, providing information about both the amplitude and the phase of the basilar membrane motion. Therefore replacing the envelope detector with a model of inner hair cell behaviour may improve sound perception.

A simple inner hair cell model (§2.4.2) can be obtained by making a small change to the HWR envelope detector (Figure 4.6c): the low-pass smoothing filter can have its cut-off frequency increased to approximately 1000 Hz, or the filter can be omitted entirely. In the Phase-Locked CIS strategy (Peeters *et al.* 1993), the two lowest frequency channels were half-wave rectified. The remaining channels had a standard envelope detector. Anecdotal results for one Laura cochlear implant subject were reported, claiming better low-frequency pitch perception, and better discrimination of male and female voices. This strategy was later implemented in the Laura Flex processor from Antwerp Bionic Systems, and made available as an option in their clinical software.
The HiRes strategy from Advanced Bionics also uses half-wave rectification (Wilson et al. 2004; Wilson et al. 2005), as can be seen in Figure 6.1, taken from an Advanced Bionics white paper (Firszt 2004). In a clinical trial (Koch et al. 2004), 30 newly-implanted recipients had three months experience with "conventional" strategies (either CIS, SAS or PPS) and then were switched over to the HiRes strategy. Unfortunately, no conclusions can be made regarding the benefit of HWR, because of the other confounding variables: the conventional strategies used 8 channels, while HiRes used 16 channels and higher stimulation rates. In a large study involving a variety of cochlear implant devices (Gfeller et al. 2007), there was no difference in pitch ranking ability between the 18 subjects who used HiRes and the 82 subjects who used envelope-based strategies (CIS, SPEAK, and ACE).

Within-subject comparisons of HWR and envelope detection will be described in chapters 10 and 12.

![Figure 6.1 HiRes processing (Firszt 2004). Top: band-pass filter output; Middle: half-wave rectifier output; Bottom: pulse sequence.](image)

### 6.2.2 Half-wave gating (HWG)

In a pilot study for this thesis using pure tones with HWR, one subject described a "vibrato" percept at certain frequencies, which appeared to be due to an amplitude modulation artefact. To avoid this artefact, the Half-Wave Gating (HWG) strategy was developed. It was described in a 2002 patent application by Swanson et al. (Table 6.1). HWG on-off modulated the quadrature envelope according to the sign of the in-phase filter output, as shown in Figure 6.2. The intent was to retain the fine timing information of HWR, whilst providing a steady amplitude during
the "on-time" so that sampling did not introduce unwanted amplitude variations. As the duty cycle was 50%, the total number of pulses per second was independent of the modulating frequency, and so loudness was expected to be relatively constant with changes in the modulating frequency (§3.7) (Tong et al. 1983a; Busby and Clark 1997).

![Figure 6.2 Half Wave Gating (HWG)](image)

### 6.2.3 F0-CIS

Geurts and Wouters (2001) evaluated pitch perception with three strategies (Figure 6.3): standard CIS, which used full-wave rectification and a 400 Hz low-pass smoothing filter; "Flat-CIS", in which the smoothing filter cut-off was reduced to 50 Hz to remove amplitude modulation cues; and a new strategy denoted F0-CIS, which had three features designed to enhance amplitude modulation cues. Firstly, F0-CIS used a zero-phase filterbank, so that F0 modulations in each channel were in phase. Secondly, it used half-wave rectification to better emulate the phase-locked behaviour of auditory nerve fibres (as in §6.2.1). Thirdly, the depth of F0 modulation was increased by subtracting an attenuated, smoothed version of the envelope.

Four Laura cochlear implant recipients pitch-ranked synthetic vowels (/a/ and /i/) at base frequencies of 150 and 250 Hz. No significant difference was found between the F0-CIS and standard CIS strategies. However, both of these strategies gave significantly better performance than the Flat-CIS strategy, at least verifying that the amplitude modulation cue was being used. One criticism of this study is that an adaptive 2-down 1-up procedure (§2.3.3) was used to measure the frequency difference limen. There were several instances of the procedure failing to converge. As mentioned in §5.2.1, such procedures assume that the underlying psychometric function is monotonic, which is not always true.
6.2.4 Modulation Depth Enhancement

The Modulation Depth Enhancement (MDE) strategy was described in a 2003 patent application by Vandali and Van Hoesel (Table 6.1) and implemented on the SPEAR research processor (Vandali et al. 2005). Each channel was processed independently, as shown in Figure 6.4. A 10 ms sliding time window was used to find the peaks and troughs in the envelope. If the calculated modulation depth was less than 6 dB, it was enhanced by reducing the amplitude in the troughs, while preserving the peak amplitude. There was no significant difference between ACE and MDE in the group mean pitch ranking score with sung vowels for five subjects (Vandali et al. 2005, study 1).

An additional optional step is to adjust the delay on each channel to align the modulation patterns in time.
6.2.5 Cochlear model strategies

This PhD thesis originated in a collaborative research project between Cochlear Ltd and the University of Melbourne to apply a cochlear model to the sound processing for the Nucleus 24 implant. The project resulted in the "Travelling Wave" patent application filed in 2000 by Blamey, Swanson et al. (Table 6.1). It was hoped that some important aspects of the excitation pattern that were clearly missing from the standard ACE strategy could be reproduced. The implementation was based on the FFT vector-sum filterbank described in §4.5.1, running at an analysis rate of 16000 FFTs per second (which precluded a real-time implementation). The weighting matrix in Equation 4.21 was selected to provide filter magnitude responses with sharp high frequency roll-offs and shallow low frequency roll-offs, similar to those shown in Figure 2.8. Each channel was delayed to match the travelling wave delays in a normal-hearing cochlea, as measured by Donaldson and Ruth (1993). Inner hair cell behaviour was modelled by a half wave rectifier (as in §6.2.1). The resulting excitation pattern was then sampled in a similar manner to ACE. The scope of the present author's work within the larger project focussed on the effect of HWR as an alternative to envelope detection (to be reported in chapters 10 and 12). Some experiments with the Lyon cochlear model filterbank will be reported in §10.3.

Wilson, Lawson et al. (2003) described their plans for the use of a cochlear model in cochlear implant processing. Pitch perception was not mentioned. They proposed to use a dual resonance nonlinear (DRNL) filterbank (Meddis et al. 2001), and the Meddis (1988) hair cell transduction model. The model's probability of spike emission would control the current of a sequence of interleaved pulses. In a subsequent NIH progress report (Schatzer et al. 2003), consonant perception was tested with a DRNL filterbank and standard FWR envelope detection (§4.4.2). No improvement was seen over CIS.

6.3 Within-channel pulse timing cues

The strategies in this category use a bank of band-pass filters and control pulse timing on each channel independently. Block diagrams of the processing in each channel are shown in Figure 6.5, with CIS included for comparison; the amplitude mapping stage is not shown. If this approach is to be implemented with sequential stimulation, then the timing of pulses on each channel cannot be truly independent: either the sampling scheme must explicitly interleave the channels, or else an arbitration scheme is required to either delay or discard pulses to resolve timing clashes.
Figure 6.5 Within-channels pulse timing strategies
6.3.1 Peak Picker

In this approach, each band-pass filter output is monitored and a pulse is produced only when it reaches a temporal peak. This provides a form of phase-locking. The earliest example of this approach was described by Wilson and colleagues at Research Triangle Institute (RTI) in their National Institute of Health (NIH) progress reports (Wilson et al. 1990; Wilson et al. 1991b; Wilson et al. 1992). Their strategy used the same front end and six-channel filterbank as the CIS strategy that had been used with Ineraid implant recipients. To maintain the interleaving of pulses, channels were allocated to consecutive time slots in a round-robin manner. No envelope detector was used. Instead, if the corresponding filter output waveform had exhibited a temporal peak since the previous time slot for that channel, then a pulse was emitted; otherwise the time slot was left empty. The filter output peak amplitude was mapped to current in the same manner that envelope amplitude was mapped in CIS. Wilson et al. called this the Peak Picker strategy, a rather unfortunate choice of name because it could potentially be confused with the concept of picking spectral peaks (as in SPEAK). Perhaps a better name would be the Temporal Peak Picker strategy. An equivalent block diagram is shown in Figure 6.5b. An issue that arises is that if a temporal peak occurs on a channel just after the time slot for that channel, then the pulse is delayed until the next time slot. Thus the pulse timing quantisation is $N$ time slots, where $N$ is the number of channels. For example, if implemented for the Nucleus 24 implant with $70 \mu s$ time slots, a 6-channel strategy would have a timing quantisation of $420 \mu s$, but a 22-channel strategy would have a quantisation of $1540 \mu s$. An experiment was conducted in this thesis to investigate the perceptual effect of such quantisation (§9.4).

The Peak Picker strategy was evaluated with a single subject. Compared to a CIS strategy with matching parameters, consonant scores were similar, and vowel scores were improved, but the differences were small because of ceiling effects (Wilson et al. 1990). A hybrid of the CIS and Peak Picker strategies (denoted PP/CIS) was later trialled with the same subject (Wilson et al. 1991b). In this strategy, temporal peak sampling was used on the two most apical channels, and the remaining four channels had the envelope detectors of CIS. The subject scored at or near 100% in speech perception tests with both PP/CIS and with his standard CIS strategy. No pitch perception testing was performed, but the subject reported anecdotally that music sounded "wonderful" with PP/CIS. Despite these promising results, the strategy was apparently not tested with other subjects, and it was not mentioned in the peer-reviewed publications by the RTI researchers.

6.3.2 Peak Derived Timing

The temporal peak sampling approach was also used in the Peak Derived Timing (PDT) strategy (Figure 6.5c), described in a 2001 patent application by van Hoesel (Table 6.1). Each channel was sampled when a peak was detected, and then any timing clashes were resolved in a
subsequent pulse arbitration stage. When multiple channels had a peak in the same time slot, priority was given to the channel with lowest frequency (thus pulse timing on the most apical channel was quantised to one time slot). For the higher frequency channels, the amplitude also affected priority. To limit the stimulation rate on each channel, a peak was ignored if it occurred within a minimum time after a previous peak.

PDT was first used in a study with bilateral implants (van Hoesel and Tyler 2003). By aligning stimulation pulses with the temporal peaks of the filter outputs, interaural timing differences could be preserved, without needing a connection between the left and right processors. In the initial implementation, only 10 channels were used due to computational limits. In a later pitch perception study with sung vowels (Vandali et al. 2005), the number of channels was increased to 19 (ideally all available channels would be used). There was no significant difference between ACE and PDT in the group mean pitch ranking score of five subjects (Vandali et al. 2005, study 1).

6.3.3 Multirate

The Multirate strategy was disclosed in a 1999 patent application by Wolfe, Carter et al. (Table 6.1). In the simplest embodiment, a pulse was requested on a channel each time that there was a positive-going zero-crossing on the corresponding filter output signal (Figure 6.5d). This approach is very similar to temporal peak sampling, because the sequence of positive-going zero-crossings has essentially the same timing as the sequence of temporal peaks (except for a 90° phase shift), as long as the filter bandwidth is relatively narrow. Compared to temporal peak sampling, one disadvantage of using zero-crossings is that the amplitude of the signal must be obtained separately. In the Multirate patent, this was achieved by having a separate envelope detector for each filter, as in CIS. If multiple channels requested a pulse at the same time, the pulse arbitration scheme gave priority to the channel having largest amplitude.

The Multirate patent further generalised the scheme to use a period estimator (Figure 6.5e). A suggested embodiment was to calculate the time between zero-crossings, then apply a smoothing filter to the series of periods. Whenever a pulse was emitted, the smoothed period estimate was used to set a timer to trigger the next pulse. By this means, the pulse rate on each channel was equal to the short-time mean frequency of the corresponding filter output signal. The pulse rate varied over time, according to the audio signal. The average pulse-rate on a particular channel was equal to the filter's centre frequency (hence the name Multirate). The maximum pulse rate on any channel was limited, to say 1000 pps. One drawback of this method was that the pulses were no longer phase-locked to the filter output signal, and the relative timing of the pulses on adjacent channels was not controlled.

Several versions of the Multirate strategy were implemented on the SPEAR research processor
by Fearn (2001, chapter 9). The ADC sample rate was 16 kHz. Initially, the filterbank was implemented as 22 fourth-order IIR filters, but the heavy computational load gave an unacceptably short battery life of 2 hours. Therefore, the IIR filterbank was replaced by a 128-point FFT filterbank (as described in §4.5), with an analysis rate of 1000 Hz (i.e. one new output of each complex band-pass filter for every 16 input samples). The amplitude of a pulse was given by the quadrature envelope (Equation 4.10). The five channels that had centre frequencies below 1000 Hz had variable pulse rates. The remaining 17 channels had a fixed pulse rate of 1000 pps. In the low frequency channels, the zero-crossing scheme had to be abandoned, as the filter signals were not sampled often enough. Instead, as illustrated in Figure 6.5f, the phase \( \phi(k, n) \) of the \( k \)th channel was calculated from the real part \( x(k, n) \), and imaginary part \( y(k, n) \) by:

\[
\phi(k, n) = \arctan \left( \frac{y(k, n)}{x(k, n)} \right) \tag{6.1}
\]

Then the instantaneous frequency \( f(k, n) \) in each channel was estimated from the change in phase between successive outputs:

\[
f(k, n) = \frac{\phi(k, n) - \phi(k, n - 1)}{T} \tag{6.2}
\]

where \( T \) is the sampling period. Finally, the reciprocal of the instantaneous frequency was used to set the stimulation period in each channel. This implementation was denoted Voc-L. Again, a drawback of this scheme was its lack of phase-locking.

Another variant of the Multirate strategy was denoted Instrument-L. To improve the frequency resolution, the ADC sample rate was reduced to 12 kHz, which reduced the FFT bin spacing from 125 to 94 Hz. This gave seven frequency bands with centre frequencies below 1000 Hz. The lowest three frequency bands were each allocated to two stimulation channels. If the estimated instantaneous frequency was in the lower half of the frequency band, then the more apical channel was stimulated, otherwise the more basal channel was stimulated. Once again, phase-locking was absent.

Pilot testing with two subjects was reported by Fearn (2001, chapter 10) using harmonic tones over a three octave range (146 – 1046 Hz). Regrettably, an adaptive pitch-ranking procedure was used, and scores varied greatly from session to session, perhaps due to a lack of convergence (§5.2.1). Compared to ACE, one subject scored better with Voc-L and Instrument-L, but the other scored worse.
6.3.4 Spike-based Temporal Auditory Representation (STAR)

The STAR strategy, described in a 2002 patent application by Grayden et al. (Table 6.1), had similarities to Peak Derived Timing (Figure 6.5c) and the zero-crossing-based Multirate strategy (Figure 6.5d). In STAR, a pulse was requested when the filter output signal crossed a threshold. The amplitude of the pulse was determined by the subsequent peak of the filter output signal. The novel aspect was that the threshold in each channel could be continuously adjusted based on previous activity in that channel, modelling neural adaptation.

6.3.5 Asynchronous Interleaved Sampling

Another approach for conveying temporal cues was proposed by Sit, Simonson et al. (2007). The processing for each channel was split into an envelope path and a phase path, as shown in Figure 6.5g. The envelope detector used a full-wave rectifier, followed by a peak detector, similar to that in §4.4.1, but with a 1 ms attack time and a 3 ms decay time. The phase path consisted of a half-wave rectifier, followed by an integrate-and-fire neuron circuit.

The main criticism of the study by Sit et al. (2007) is that it only used normal hearing subjects, listening to an audio reconstruction. Any claims that cochlear implant recipients would derive benefit from this processing strategy are pure speculation.

6.3.6 Channel-specific sampling sequences (CSSS)

This strategy was described in a 1999 patent application by Zierhofer (Table 6.1). The MED-EL Pulsar implant has the ability to stimulate on multiple electrodes simultaneously. One of the challenges with simultaneous stimulation is transmitting sufficient information over the RF link to specify all of the pulses. Zierhofer's solution is to store in the implant a short sequence of pulses for each electrode. In one method, the amplitudes of each sequence are samples of a half-cycle of a sinusoid at the centre frequency of the corresponding filter. In the sound processor, the output of each filter is half-wave rectified and compressed. On each positive zero-crossing, the processor transmits the peak amplitude of the previous half-wave segment. The implant emits the stored sequence of pulses on the corresponding electrode, scaled according to the received amplitude value. Multiple channels can emit their sequences concurrently. If the bandpass filter is relatively narrow in bandwidth, the pulses are a reasonable approximation to those produced by a strategy using HWR (§6.2.1). The strategy also has aspects of the pulse timing approach because an amplitude is only sent when a zero crossing occurs, which provides the beneficial reduction in the RF link data transmission rate. Like other strategies using the pulse timing approach, the scheme faces timing quantisation and arbitration issues (only one zero-crossing event can be sent across the link at a time).

MED-EL's Fine Structure Processing (FSP) strategy uses CSSS on the two most apical
Improving cochlear implant pitch perception

channels, and standard CIS-type processing on the remaining channels (Hochmair et al. 2006). Arnoldner et al. (2007) switched 14 recipients from CIS on the Tempo+ processor to FSP on the Opus processor. They used a music test battery comprising three sub-tests: melody discrimination, rhythm discrimination, and a "number of instruments" test (which asked how many different instruments a subject could distinguish in a short piece of music). Melody discrimination is clearly the most relevant to pitch perception, yet subjects did significantly worse with FSP than with CIS. There was no significant difference in speech perception scores between strategies at the initial switch-over. Subjects' speech scores did improve over the next three months, but as the study design was not balanced, it was not possible to conclude whether the improvement was due to the change in strategy, the change in processor hardware, or the usual progress with implant experience (the subjects averaged less than one year of implant use at study commencement).

In a conference poster, Mitterbacher et al. (2005) reported better harmonic tone pitch perception with CSSS than with CIS in three subjects. Surprisingly, performance was about the same for fundamental frequencies of 220 and 440 Hz (temporal cues are expected to be less effective at 440 Hz). One criticism is that the experiment used an adaptive 4AFC discrimination task, so it is possible that subjects used non-pitch cues, or there may have been convergence issues (§ 5.2.1).

In a conference presentation, Brill et al. (2007) described a multi-centre study with 46 recipients; FSP was reported to give an improvement in speech understanding compared to CIS, and a greater range in numerical pitch estimation (§ 2.3.4) for pure tones. In a pre-symposium MED-EL workshop, Au (2007) described an implementation of FSP in which a single pulse was emitted on each zero crossing on the two most apical channels. This is similar to Wilson's PP/CIS (§ 6.3.1).
Figure 6.6 Global temporal cue strategies
6.4 Global temporal cues

The strategies in this section convey a "global" temporal cue to the fundamental frequency on all the channels synchronously. With reference to Table 6.2, this can be done in two ways: either by varying the pulse rate, or by applying amplitude modulation to fixed high-rate pulse trains. Representative block diagrams are shown in Figure 6.6.

6.4.1 F0 pulse rate

In this approach, the fundamental frequency (F0) of the audio signal is estimated, and used to control the rate of stimulation, taking advantage of rate-pitch cues (Figure 6.6a). Although the pulse rate varies with time, at any moment a common (i.e. "global") pulse rate is used for all active channels. This simplifies the interleaving of pulses across channels.

The Nucleus 22 system initially used this approach. Successive refinements were made over the years, but the common principle was that F0 was estimated, and in each fundamental period, a subset of electrodes was stimulated. The MPEAK strategy selected four electrodes each period (there was insufficient time for any more pulses due to the pulse widths in use). The pulse amplitudes were determined by the envelopes of band-pass filters. For unvoiced sounds, a quasi-random rate (averaging 250 Hz) was used.

This approach went out of favour with the introduction of the SPEAK strategy in 1994 because many studies showed that SPEAK improved speech perception compared to the F0-based strategies. However, surprisingly little research was conducted into pitch and music perception, perhaps because the goal of good speech perception seemed challenging enough at that point in history.

Interest in this type of strategy has recently been revived. Green et al. (2004) tested a strategy using an 8-channel filterbank, where F0 was estimated and a single pulse was output on each channel each fundamental period, so that F0 directly controlled the pulse rate. This strategy will be further described in §6.4.2.

This type of strategy was also included in experiment §10.3 of this thesis.

6.4.2 F0 Modulation

An alternative means of using a fundamental frequency estimate is to generate a signal having this frequency and then use it to modulate the envelopes of all of the channels (Figure 6.6b). The benefits of this approach are that all channels modulate in phase, and the modulation waveform and depth can be controlled. There are several challenges, all of which were also encountered with the early Nucleus 22 F0 strategies. It is difficult to implement a robust F0
estimator, especially at poor signal-to-noise ratios, or when there are competing voices. Modulation should not be applied if the sound is not periodic (e.g. unvoiced speech, or noise). It is not obvious what to do if there are multiple F0s present (e.g. two competing voices).

The envelope detector must be designed to suppress the natural F0 modulation of the filter envelopes, i.e. it should have a low-pass cut-off frequency below the lowest expected F0, unlike the usual 200 – 400 Hz cut-off used in CIS or ACE.

Several research groups appear to have developed this idea independently.

The F0Sync strategy was described in a 2003 patent application by McDermott and McKay (Table 6.1). The proposed F0 estimator used a 128-point FFT filterbank and looked at the rate of change of phase. An estimate of the ratio of harmonic-to-inharmonic power was used to decide if modulation should be applied. A subsequent implementation on the SPEAR research processor (Vandali et al. 2005) used a simpler F0 estimator: the audio signal was low-pass filtered at 450 Hz, peak detected and differentiated, and finally the time intervals between zero-crossings were measured. Envelopes were smoothed by a 50 Hz low-pass filter. The modulating signal was a square wave (i.e. on-off modulation) with 33% duty cycle (Figure 6.7a). The group mean score of three subjects for sung vowel pitch ranking was significantly higher with F0Sync than ACE (Vandali et al. 2005, study 2). No speech perception results with F0Sync were reported.

A similar strategy was investigated at University College, London (Green et al. 2004). It was based on the eight-channel CIS strategy used with the Clarion implant. Smoothed envelopes were obtained by using a full-wave rectifier followed by a 32 Hz low-pass filter. The study used synthesised speech stimuli, with known F0s, so an F0 estimator was not required. Two modulating waveform shapes were trialled: sawtooth (Figure 6.7b), and a temporally sharpened "sawsharp" waveform (Figure 6.7c). As mentioned in §6.4.1, a condition was also trialled where only a single pulse was produced per fundamental period (Figure 6.7d, i.e. SPP in the terminology defined in §3.7). In a pitch-glide labelling experiment with eight Clarion recipients, the F0-modulated conditions gave better performance than standard CIS (the improvement was statistically significant, but small). The sawsharp and SPP conditions had very similar performance. This is consistent with the psychophysical results reviewed in §3.7 (e.g. McDermott and McKay 1997), where both forms of temporal pitch (modulation frequency and pulse rate) gave similar percepts. Green et al. hypothesised that the SPP and sawsharp conditions were more likely to produce the desired inter-spike intervals in the neural firing pattern. As expected, the benefit diminished when F0 exceeded 250 Hz. Unfortunately, in a subsequent study (Green et al. 2005), vowel perception was found to have been degraded compared to CIS. This was unexpected, as spectral information should still be well represented in the stimulation pattern. A possible cause is that the sawsharp modulating waveform only
The F0mod strategy was developed in Leuven (Laneau 2005, chapter 7; Laneau et al. 2006), and initially implemented in MATLAB, using the Nucleus MATLAB Toolbox implementation of ACE (§7) as a starting point. F0mod used an FFT filterbank, as described in §4.5, except that the FFT length was increased to 512 points. This reduced the FFT frequency spacing to 31.25 Hz, and (more importantly) reduced the bandwidth of each FFT bin filter to 62.5 Hz. The power-sum method was used to combine the 256 narrow bin filters into 22 channels. Thus the quadrature envelopes were inherently band-limited to suppress F0 modulations, without needing any additional low-pass filtering. The analysis rate was 1778 Hz (one FFT every nine input samples). The F0 estimate was obtained by calculating the magnitude squared of each FFT bin (i.e. the power spectrum), then performing an inverse FFT, to give the circular auto-correlation. The lag with the highest auto-correlation value gave the fundamental period. Sinusoidal modulation was applied with 100% depth. Subsequent processing (maxima selection, amplitude compression, and current level scaling) was identical to ACE. The F0mod strategy was evaluated with six Nucleus 24 recipients. Pitch-ranking was performed using the method of constant stimuli with intervals of 1, 2, and 4 semitones. The stimuli were MIDI-synthesised instruments (piano, clarinet, trumpet, guitar, and voice). The group mean percent-correct scores were significantly higher for F0Mod than ACE at base frequencies of 131 and 185 Hz (C3 and F#3). There was no significant difference between the two strategies at 370 Hz (F#4), which was to be expected, as temporal cues were unlikely to be salient at that frequency (§3.7). Four of the subjects performed closed-set familiar melody identification using ten isochronous melodies (§5.2.3) played on the MIDI clarinet, and obtained significantly better scores with F0Mod than with ACE. Speech perception was not tested.

A strategy using the F0 Modulation approach was included in the experiment described in §10.3 of this thesis. Several different modulating waveforms were investigated.
6.4.3 Multi-channel Envelope Modulation (MEM)

The Multi-channel Envelope Modulation (MEM) strategy was disclosed in a 2004 patent application by Vandali, van Hoesel and Seligman (Table 6.1). As shown in Figure 6.6c, this strategy was similar to F0 Modulation, the difference being that an explicit F0 estimator was not required. Instead, a modulating signal was generated directly from the audio input, by rectifying and filtering it. It was implemented on the SPEAR research processor, and compared to other strategies in a sung vowel pitch ranking study (Vandali et al. 2005, study 3). Four out of five subjects showed a statistically significant improvement with MEM compared to ACE, and the group mean improvement was statistically significant, although small (ACE 73%, MEM 80%). Regarding speech perception, there was no significant difference between MEM and ACE for CNC words or CUNY sentences in noise.

6.4.4 F0 modulates pulse rate

Another strategy using an F0 estimator to control the global pulse rate was proposed by Lan, Nie et al. (2004). Instead of the pulse rate being constant (as in CIS), or being equal to F0 (as in the "Single" condition of Green et al. (2004)), the pulse rate $R$ was modulated up and down by the estimated value of F0, i.e.:

$$ R = R_0 + (F_0 - 200) $$

where $R_0$ was the base pulse rate, and $F_0$ was the estimated fundamental frequency, which had an expected range of approximately 100 to 300 Hz. The base pulse rate was not given by Lan, Nie et al. (2004), but was apparently intended to be more than 800 pps, as is typical for CIS. This approach seems to be based on the questionable assumption that place and temporal cues are interchangeable (§3.8.2), and that the pitch of an electrical pulse train on a specific electrode can be shifted up and down by varying the pulse rate. Using pulse rates beyond 300 pps appears misguided, as pulse rate variations beyond this rate are not perceived as pitch changes (§3.7). There are no published reports of this strategy being tested with any cochlear implant recipients.

6.5 Enhanced place cues

Sound coding strategies in this category attempt to enhance place cues to pitch.

6.5.1 Virtual channels

Wilson and colleagues developed the CIS strategy for the Ineraid implant, which had six
intracochlear electrodes. They found that speech perception scores decreased monotonically as the number of channels was reduced from six (Wilson et al. 1994). Conversely, it suggested the possibility of better performance with more channels. As described in §3.6.1, when two adjacent monopolar channels are simultaneously stimulated, an intermediate place-pitch percept can be obtained. Using this technique, they formed a virtual channel between each pair of electrodes, thus increasing the number of channels to eleven (six "physical" and five virtual channels). An 11-channel CIS strategy (denoted "VCIS") could then be used. In three Ineraid subjects, speech perception scores for VCIS were similar to or less than scores with standard CIS (Wilson et al. 1995). An extensive study of 16 different configurations of physical and virtual channels was conducted with one subject. Regression analysis showed that 83% of the variation in consonant scores was accounted for by the total intracochlear distance spanned by the channels; inserting virtual channels had no significant effect (Wilson et al. 1995). Wilson et al. acknowledged that the lack of benefit for virtual channels may have been because intermediate place-pitch percepts were already being produced by sequential stimulation in standard CIS (McDermott and McKay 1994). Despite this, subjects expressed a preference for VCIS, and reported that music sounded richer and more natural (Wilson et al. 1995), although no formal pitch perception testing was performed.

6.5.2 Current Steering

The Advanced Bionics HiRes120 strategy is based on the current steering principle: two adjacent electrodes are pulsed simultaneously. As the ratio of the currents between the two electrodes is varied, the effective place of stimulation can be steered between the two electrodes (§3.6.1). For commercial reasons, few technical details have been published.

The HiRes120 strategy is similar to CIS, except that each filter band is allocated to the region between two adjacent electrodes, rather than to an electrode itself. The HiRes90K implant has 16 intracochlear electrodes, thus the audio signal is divided into 15 frequency bands. In each analysis period, 15 pulses are delivered (Buechner et al. 2008). The filter envelope determines the pulse amplitude. Each pulse stimulates two adjacent electrodes, with the ratio of the currents on the two electrodes determined by a "current navigator", as shown in Figure 6.8 (Advanced Bionics Corporation 2006). Note that this is quite different from the virtual channel CIS strategy (Wilson et al. 1995) described in §6.5.1, which would use 31 filters for 16 intra-cochlear electrodes.

One implementation (Figure 6.9) was disclosed in a 2002 patent application by Litvak et al. (Table 6.1). An estimate of the instantaneous frequency $F_i$ in each band was obtained by counting the number of zero crossings in a 20 ms time window. The current steering factor $\alpha$ (the proportion of current on the more basal electrode, as in §3.6.1) was then calculated by linear interpolation (on a logarithmic scale) between the frequencies allocated to the two
corresponding electrodes, i.e.:

\[
\alpha = \frac{\log F_H - \log F_L}{\log F_H - \log F_L}
\]  

(6.4)

where \(F_L\) is the lower frequency and \(F_H\) is the higher frequency. Although \(\alpha\) could in principle be calculated to any desired precision, in the HiRes120 strategy it is apparently quantised to one of 8 possible values. Thus it is claimed that a pulse can be delivered to any one of \(8 \times 15 = 120\) "sites". The excitation pattern produced is broader than that produced by monopolar stimulation on a single electrode, and each of the possible patterns heavily overlaps with its neighbours. A more accurate description would be that the broad excitation pattern produced by each pair of electrodes has 8 possible centroids.

Figure 6.8 HiRes strategy block diagram (Advanced Bionics Corporation 2006)

Figure 6.9 Current steering strategy – processing for one filter band.
The HiRes120 clinical trial apparently used a different implementation, based on a 256-point FFT (Advanced Bionics Corporation 2007). In the multi-centre clinical trial, experienced recipients were switched from the older HiRes strategy (§6.2.1) running on an older processor, to the new HiRes120 strategy running on the new Harmony processor, which had an improved front end. Thus it was not possible to judge whether performance changes were due to the strategy or the processor hardware (Brendel et al. 2008). An extended protocol at the Medical University of Hannover with 9 recipients removed this confounding processor variable by including an additional phase using the older HiRes strategy on the new Harmony processor (Brendel et al. 2008). There was no significant difference in scores between the HiRes and HiRes120 strategies (on the same processor hardware) for sentences in quiet, or in speech-shaped noise, or with a single competing talker. For sentences with a competing talker, the older hardware gave significantly worse scores than either strategy on the newer hardware. Pitch perception was not measured.

Another Hannover study (Buechner et al. 2008) compared HiRes to three different current steering strategy implementations, denoted SpecRes120, SpecRes16k, and SinEx. The body-worn Platinum sound processor was used for all four strategies. Eight recipients participated in the study; each had two phases of one month of use with each strategy. There was no significant difference between the four strategies in the group mean percent correct scores for sentences (in quiet, in speech-shaped noise, or with a single competing talker); or in sound quality ratings. Pitch perception was not reported. Few details on the signal processing were provided. The SinEx strategy (also described in Nogueira et al. 2007) decomposed the audio signal into a sum of (at most) 15 sinusoidal components. SpecRes120 was described as identical to HiRes120, using an FFT analysis with 15 frequency bands; within each band, the current steering factor $\alpha$ was determined by the "highest spectral peak" in the band. SpecRes16k used the same FFT analysis, but was claimed to provide 16000 stimulation sites. Based on the sketchy information available, some educated guesses can be made concerning implementation. Assuming an audio sampling rate of 16 kHz, a 256-point FFT yields 128 frequency bands (§4.5), with a linear spacing of 62.5 Hz. The 128 FFT bins were presumably grouped into 15 frequency bands in a similar manner to the FFT filterbank described in §4.5, using an approximately logarithmic frequency scale. In that case, the lowest frequency bands would be comprised of only two or three FFT bins. Taking the position of the largest bin would only provide two or three different current steering factors ($\alpha$) per frequency band. A plausible alternative approach would be to take the centroid of the FFT bin magnitudes (using Equation 3.7, where each $a(k)$ would be an FFT bin magnitude), which could be calculated to any desired precision: 3 bits gives 8 possible values of $\alpha$ per band (120 "sites"); 10 bits gives 1024 possible values of $\alpha$ per band (almost 16000 "sites").

Does current steering improve spectral resolution? Figure 6.10a depicts the acoustic excitation
Improving cochlear implant pitch perception

6.6 Residual acoustic hearing

An entirely different approach to improving the pitch perception of cochlear implant recipients is to focus not on electrical stimulation, but on their residual acoustic hearing. The candidacy
criterion for a cochlear implant in 1985 was profound hearing loss in both ears. However, as the technology and the outcomes have improved, the criteria have been broadened and many recipients today have some usable hearing, predominantly at low frequencies.

6.6.1 Bimodal

Using a cochlear implant in one ear and a hearing aid in the opposite (contralateral) ear is known as bimodal hearing. Even though the hearing aid by itself may provide little or no speech recognition, speech perception in the bimodal condition may still be better than with a cochlear implant alone.

Kong, Stickney and Zeng (2005) tested cochlear implant recipients under three conditions: cochlear implant alone, contralateral hearing aid alone, and bimodal. They used a closed-set familiar melody identification task, utilising 12 melodies with rhythm cues removed (§5.2.3). Table 6.3 shows the raw scores of the four regular users of contralateral hearing aids (S1, S3, S4, and S5), graciously provided by Dr Kong. Columns 3, 4, and 5 of the table show the number correct out of 36 trials for the three conditions. For each subject, the row labelled "Low" shows scores for melodies in the frequency range 104 – 262 Hz (G#2 – C4). Similarly, "Mid" labels scores for melodies one octave higher (G#3 – C5), and "High" labels scores for melodies two octaves higher (G#4 – C6). The row labelled "mean" shows the mean score across the three frequency ranges. Column 6 shows the bimodal score minus the CI-alone score. The row labelled "p-value" shows the probability of obtaining such an extreme difference in scores by chance (analysed by the present author according to the Monte Carlo method in §16.4.3). Many CI scores were near chance (3/36). For each subject, the bimodal condition gave significantly higher scores than a cochlear implant alone. In contrast, the last column compares the bimodal scores to the hearing aid alone scores, showing that those two conditions were not significantly different in any subject. This implies that in the bimodal condition, the subjects were not combining pitch information from the two ears, but were instead primarily attending to the better (hearing aid) ear.
Improving cochlear implant pitch perception

<table>
<thead>
<tr>
<th>Subject</th>
<th>Freq range</th>
<th>Bi-modal score</th>
<th>CI score</th>
<th>HA score</th>
<th>Bi - CI score</th>
<th>Bi - HA score</th>
</tr>
</thead>
<tbody>
<tr>
<td>51</td>
<td>Low</td>
<td>32</td>
<td>3</td>
<td>31</td>
<td>29</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Mid</td>
<td>33</td>
<td>4</td>
<td>32</td>
<td>29</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>31</td>
<td>21</td>
<td>32</td>
<td>10</td>
<td>-1</td>
</tr>
<tr>
<td></td>
<td>mean</td>
<td>32.0</td>
<td>9.3</td>
<td>31.7</td>
<td>22.7</td>
<td>0.3</td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>53</td>
<td>Low</td>
<td>11</td>
<td>10</td>
<td>14</td>
<td>1</td>
<td>-3</td>
</tr>
<tr>
<td></td>
<td>Mid</td>
<td>16</td>
<td>14</td>
<td>15</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>20</td>
<td>10</td>
<td>22</td>
<td>10</td>
<td>-2</td>
</tr>
<tr>
<td></td>
<td>mean</td>
<td>15.7</td>
<td>11.3</td>
<td>17.0</td>
<td>4.3</td>
<td>-1.3</td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>54</td>
<td>Low</td>
<td>10</td>
<td>4</td>
<td>8</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Mid</td>
<td>9</td>
<td>4</td>
<td>11</td>
<td>5</td>
<td>-2</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>8</td>
<td>3</td>
<td>7</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>mean</td>
<td>9.0</td>
<td>3.7</td>
<td>8.7</td>
<td>5.3</td>
<td>0.3</td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>55</td>
<td>Low</td>
<td>19</td>
<td>9</td>
<td>14</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Mid</td>
<td>18</td>
<td>19</td>
<td>17</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>16</td>
<td>12</td>
<td>14</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>mean</td>
<td>17.7</td>
<td>13.3</td>
<td>15.0</td>
<td>4.3</td>
<td>2.7</td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6.3 Familiar melody identification scores from Kong, Stickney and Zeng (2005). Scores are the number correct out of 36 trials. Chance performance is 3/36.

The possible pitch perception benefits of a contralateral hearing aid will be investigated in §11.5.

6.6.2 Electro-Acoustic Stimulation (EAS)

More recently, advances in surgical techniques and the development of new electrode arrays have allowed the preservation of low frequency hearing in the same ear as a cochlear implant. Using a cochlear implant and a hearing aid in the same ear is known as electro-acoustic stimulation (EAS). Such recipients may in addition use a hearing aid in the contralateral ear.

A team at the University of Iowa collaborated with Cochlear Ltd in the development and evaluation of a short electrode array, designed to preserve residual low-frequency hearing (Gantz and Turner 2004). Audio frequencies above 2000 Hz were presented by electrical stimulation, and below 2000 Hz acoustically. Pitch perception was measured in three groups of subjects: 13 short-electrode recipients (in the EAS condition); 101 cochlear implant recipients having a traditional "long electrode", and 21 normal-hearing subjects (Gfeller et al. 2007). Group mean scores for pure tone pitch ranking with one semitone intervals are shown in Figure
6.11. The EAS group scores were significantly greater than the long-electrode group, except at the higher frequencies (where their acoustic hearing was poorest). Although a great improvement over electrical stimulation, the EAS group still performed substantially worse than the normal hearing group. This is consistent with studies showing that hearing impairment degrades pure tone frequency discrimination (Moore and Carlyon 2005).

![Graph showing Pure tone pitch ranking group scores for Normal-hearing (NH), Acoustic + electric (A+E), and "long electrode" cochlear implant (LE) (1 semitone intervals)](graph)

Figure 6.11 Pure tone pitch ranking group scores (Gfeller et al. 2007) for Normal-hearing (NH), Acoustic + electric (A+E), and "long electrode" cochlear implant (LE) (1 semitone intervals)

6.7 Conclusion

A huge variety of novel sound processing and stimulation schemes have been proposed to improve cochlear implant pitch perception. Despite valiant efforts by many researchers, little practical benefit has been demonstrated. The only approach that has shown a substantial benefit is the use of residual acoustic hearing.
7 Nucleus MATLAB Toolbox

7.1 Introduction

MATLAB (The MathWorks Inc, Massachusetts) is an interactive mathematical programming environment that runs on most computing platforms. Nucleus MATLAB Toolbox (NMT) is a collection of MATLAB software written by the present author. It was developed to meet several goals:

- **Specification**
  MATLAB functions can act as specifications for signal processing algorithms. The traditional way of specifying an algorithm is to write a document containing material such as block-diagrams, flow charts, mathematics, and pseudo-code. It is hard to make these documents unambiguous and complete. In contrast, a MATLAB implementation is concise, because the language is high-level, and can be tested for accuracy and completeness by executing it. Within Cochlear Ltd, DSP assembly code is tested using MATLAB as a reference.

- **Visualisation**
  MATLAB provides flexible graphics, sound, and user interface capabilities that can be used to demonstrate or visualise the operation of signal processing algorithms.

- **Development**
  The MATLAB environment is ideal for developing new signal processing algorithms. Its interpreted, high-level language allows algorithms to be developed and tested much more rapidly than in DSP assembly language.

- **Experimentation**
  NMT includes functions for performing psychophysics experiments with normal hearing subjects or cochlear implant recipients. This is further described in §7.4.

7.2 Sound processing strategies

This section briefly describes the NMT implementation of the strategies used in the experimental work, with cross-references to Chapters 4 and 6. Three types of data format are used. Firstly, an audio signal is represented as a vector. The audio sampling rate is 16 kHz. Secondly, a multi-channel signal, such as the filterbank output, is represented as a matrix (most processing steps operate on an entire matrix). Time runs across the rows, and frequency runs along the columns, corresponding to the format of a spectrogram display. Thirdly, a sequential
stimulation pattern is represented as a MATLAB struct, a container with named fields, where each field is a vector (as described further below).

### 7.2.1 ACE

A block diagram of NMT ACE is shown in Figure 7.1. The NMT convention is that the names of all processing functions end with "\_proc", but for clarity this suffix is omitted here. FFT\_filterbank implements the 128-point FFT filterbank (§4.5), with a Hann window. The output is a complex-valued matrix with 65 columns, one per FFT bin. The FFT bin signals are combined into 22 real-valued envelope signals (or fewer as needed) by Power\_sum\_envelope (§4.5.2). Note that the bands below 1000 Hz are all one bin wide, so there is no difference between vector-sum and power-sum in the low frequency range. Reject\_smallest implements maxima selection (§4.6). It produces a matrix having the same size as its input, with the smallest samples in each column set to "Not a Number" (NaN), and the remaining (largest) samples left unchanged (NaN is a special value defined in IEEE floating point arithmetic that propagates through any subsequent calculations). LGF applies non-linear amplitude compression (§4.8). Up until this point the signal is represented as a matrix, to allow for spectrogram-style display. Collate\_into\_sequence scans through the matrix, skipping over the NaN values, and creates a struct with fields specifying the channel number and magnitude of each pulse in the sequence. This representation is recipient-independent. Finally, Channel\_mapping uses a particular recipient's T and C levels to produce a struct with fields specifying the active electrode, stimulation mode, current level, phase width, and phase gap of each pulse in the sequence.

In the experiments using ACE in chapter 10, the analysis rate was equal to the stimulation rate, and was an integer divisor of the ADC rate; for example a nominal 1800 pps map had an analysis rate and stimulation rate of 16000/9 = 1778 Hz, i.e. one FFT every nine audio samples.
7.2.2 Overview of research strategies

The research strategies that were used in the experimental work fall into three families: Half-Wave Rectification (HWR; §6.2.1), F0 Modulation (F0M; §6.4.1), and Temporal Peak Sampling (TPS; similar to Peak Derived Timing §6.3.2). Several variants of each approach were implemented. Table 7.1 lists the strategies used in the experiments of later chapters. Table 7.2 lists the functions used to implement these strategies. The modular nature of NMT allowed many of the functions to be re-used across several strategies.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>ACE</th>
<th>HWG</th>
<th>HWR</th>
<th>HWR Lyon</th>
<th>TPS</th>
<th>TPS Lyon</th>
<th>F0M Single</th>
<th>F0M OnOff</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pitch rank pure tones §10.2</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>Pitch rank sung vowels §10.3</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
</tbody>
</table>

Table 7.1 NMT strategies used in streaming experiments
The HWR strategy does not require a quadrature filterbank, but the FFT filterbank was used for consistency. One difference between ACE and HWR that is not apparent in Table 7.2 is the analysis rate. In HWR, the analysis rate was equal to the audio sampling rate (16 kHz); in other words, a new FFT was performed for every new audio sample. This means that the filterbank output sampling rate was equal to the audio sampling rate, just as it would be in a typical tapped delay line FIR filterbank implementation. Because the processing was performed off-line, this computational load was acceptable.

The FFT bins were combined by a vector sum to yield the complex filter outputs (Equation 4.21), and then the real part (i.e. the in-phase filter output) was half-wave rectified. A power sum cannot be used, because it only produces an envelope. The bin weighting matrix provided the same frequency boundaries as ACE.

Before the half-wave rectified filter output is down-sampled to the stimulation rate (1778 Hz), some form of smoothing is required to prevent aliasing. `Time_interval_maxima` divided the signal into successive time intervals of nine samples duration (one stimulation period, i.e. the reciprocal of the stimulation rate) and found the maximum value of each channel in each time interval. The remainder of the HWR strategy was the same as ACE.

### 7.2.4 HWG

The HWG strategy (§6.2.2) used the same FFT vector-sum filterbank as HWR, also running at an analysis rate of 16 kHz. It also used `Time_interval_maxima` to avoid aliasing.

### 7.2.5 HWR Lyon

HWR Lyon was the same as HWR, except that the FFT vector-sum filterbank was replaced with the Lyon cochlear model filterbank described in §2.4.2 (Figure 2.7). It used the implementation contained in the Auditory Toolbox, a collection of MATLAB functions written by Malcolm Slaney (Slaney 1998). The quality factor $Q$ of the filterbank was set to 8, so that the bandwidths
approximately matched the critical bands of normal hearing. The filterbank contained 88 filters. Every fourth filter output was used, giving 22 channels. The AGC shown in Figure 2.7 was not used.

7.2.6 TPS

The TPS strategy used the same FFT vector-sum filterbank as HWR. It also ran at 16 kHz, to allow the temporal peaks in the filter outputs to be located with a time resolution of one audio sample. The arbitration scheme divided the signal into successive time intervals of 9 samples duration and ensured that each channel had at most one pulse per interval. This limited the pulse rate on the high frequency channels to 1778 pps. At most 8 pulses could be delivered in this time interval (which would yield a total of 14222 pps). If more than 8 channels had a temporal peak in one interval, then the largest 8 were selected. Pulses were placed into the closest of the 8 available time slots starting with the lowest frequency.

7.2.7 TPS Lyon

TPS Lyon was the same as TPS, except that the FFT vector-sum filterbank was replaced with the Lyon cochlear model filterbank.

7.2.8 F0M

Like ACE, F0M used the FFT filterbank with power sum envelopes, and had an analysis rate equal to the stimulation rate (1778 Hz). The implementation was similar to that of Green, Faulkner et al. (2004). Because this strategy was only used in a pitch-ranking experiment with stimuli having known fundamental frequencies (§10.3), no F0 estimator was implemented. The function Collate_F0_multiple_scan used the known F0 to divide the filter envelope matrix into successive time intervals of one fundamental period in duration. In each period, the maximum amplitude of each channel was found, yielding a vector of 22 amplitudes. Then the eight channels with largest amplitudes were found by Reject_smallest, and Collate_into_sequence was used to create a sequence containing one pulse on each of those channels, representing a single "scan" across the electrode array. Collate_F0_multiple_scan implemented several variants of F0M. If the modulation_type parameter was specified as single, then a single scan sequence was emitted at the start of each period, and the remaining time-slots in the period were left empty. This can be classified as a global pulse rate cue (§6.4.1). If the modulation_type parameter was specified as on_off, then a succession of identical scan sequences were emitted at the start of each period. The number of sequences was calculated to give an approximately 50% duty cycle. This can be classified as a global amplitude modulation cue (§6.4.2).
7.3 SPrint streaming

The present author wrote the SPrint streaming software, which provides the ability to deliver a pulse sequence created in NMT to a Nucleus 24 cochlear implant recipient. A pulse sequence can be generated by processing a sampled audio signal with a sound processing strategy. Alternatively, in some psychophysics experiments (such as those in chapter 9) the fields of the sequence struct are directly specified. The software sends a data stream representing this pulse sequence via a Clinical Programming System (CPS) to a SPrint processor, which then transmits the corresponding RF signal to the implant. Each pulse is specified by its active electrode, stimulation mode, current level and phase width, which are encoded into two 16-bit words. Using the CPS, the PC can transmit 70000 16-bit words per second to the SPrint, which easily supports the Nucleus 24 implant maximum pulse rate of 14400 pps. The software that runs on the SPrint consists of two tasks. In one task, the data stream is received, decoded, and stored in a circular buffer. In the second task, information for a pulse is read from the buffer and written to the RF protocol generation hardware. Because the two tasks run concurrently, the software allows stimulation sequences of unlimited duration.

7.4 Psychophysics procedures

Nucleus MATLAB Toolbox includes a suite of functions for performing psychophysical experiments. Experiments are defined in MATLAB scripts, which call functions that construct a set of stimuli and create a stimulus presentation window. For example, Figure 7.2 shows a window for the SPP stimuli that were used in experiment §9.2. In this case, the left column shows the pulse rate of each stimulus. Each stimulus has a checkbox to enable it. Once a stimulus is enabled, it can be presented to the subject by pressing the "Stimulate" button. The sliders and edit fields allow the level of each stimulus to be adjusted. The "Sweep Up" and "Sweep Down" buttons present all of the enabled stimuli in turn.

The remaining buttons at the bottom of the window launch various psychophysical procedures (some of which are described in the following sections). Only the enabled stimuli are used in these procedures. In all procedures, a record of the stimuli presented and the subject's response is displayed on the MATLAB command window (which is hidden from the view of the subject), and also written to a log file. A complete record of the responses, as well as all the information needed to repeat the experiment, is also saved as a binary file in the MATLAB data format.
7.4.1 Loudness balancing

The Balance button (Figure 7.2) launched a Loudness Balance procedure. The two drop-down menus to the right of the "Balance" button specified the reference and the varying stimulus. The subject initiated each trial by pressing the Ready button. In each trial, a pair of stimuli was presented. In pilot studies, a simple AB sequence was used, but some subjects showed a bias towards responding that the first stimulus was louder (even if they were identical). Instead, in
the experimental work in this thesis, both stimuli were presented twice in an alternating pattern, i.e. ABAB.

![Loudness Balance Procedure](image)

**Figure 7.3 Loudness balance procedure**

The software supports a variety of adaptive rules to change the level of the varying stimulus from trial to trial, based on the subject's response. The experimental work in this thesis used the QUEST rule (§2.3.3), implemented in the MATLAB Psychophysics Toolbox by David Brainard and Denis Pelli (Brainard 1997; Pelli 1997).

### 7.4.2 Pitch ranking

The majority of the experiments in chapters 8, 9, and 10 used the same pitch ranking procedure, so it will be described in detail here. Any exceptions to this standard procedure will be pointed out in the description of that experiment.

The method of constant stimuli was used (§2.3.3). A common set of fundamental frequencies was used in several experiments. For the synthesised stimuli (pure tones and single-channel sequences), the entire set consisted of 25 frequencies, with one semitone spacing, spanning a two-octave range from 99 to 397 Hz, as shown in Figure 7.2. It covered the fundamental frequency range of most voiced speech. The sung vowel stimuli (§10.3) covered a smaller frequency range (98 – 330 Hz), but also had a one semitone spacing.
The difficulty of the pitch ranking task was controlled by the choice of frequency interval. A frequency interval of $M$ was achieved by enabling every $M$th stimulus. For example, Figure 7.2 shows the six-semitone subset. In a block of trials, a subject always ranked a pair of stimuli that were adjacent in the subset, i.e. each pair had a constant frequency interval. Intervals of 1, 2, 4, and 6 semitones were used in various experiments in this thesis.

A block consisted of eight trials of each pair, in randomised balanced order. For the synthesised stimuli, the six-semitone set contained five stimuli, i.e. four adjacent pairs, so its block comprised 32 trials. The four-semitone set contained six adjacent pairs, so its block comprised 48 trials, which typically took five minutes to perform. To avoid subject fatigue, this was the maximum number of trials presented in a block. Two-semitone and one-semitone intervals were split into multiple blocks each containing six pairs (48 trials).

Once the appropriate subset of stimuli was selected, the lowest frequency stimulus was adjusted to be medium loud, then successive adjacent pairs of stimuli were loudness-balanced (§7.4.1). Finally, all the stimuli in the subset were presented sequentially (using the "Sweep" buttons) to confirm that they were equal in loudness. The "Save" menu saved the resulting levels to a file so that they could be reloaded in any later session.

The "Pitch rank" button (Figure 7.2) was then pressed, launching the Pitch Ranking window (Figure 7.4). This window was used by the subject. It was displayed on a touch screen, which was attached to the computer as a second monitor. The subject initiated each trial by pressing the Ready button. In each trial, a pair of stimuli was presented, with the first stimulus repeated, i.e. AAB. Presenting stimulus A twice was intended to establish it as a perceptual reference to compare B against. The subject was asked whether the final sound was higher or lower in pitch than the preceding sound, and responded by clicking the appropriate button. The synthesised stimuli were 500 ms in duration and were separated by 250 ms of silence. No trial-by-trial feedback was provided. At the end of a block, subjects often enquired into their performance; if so, they were told the overall score for that block. This helped maintain their motivation.
7.4.3 4AFC

The Discriminate button (Figure 7.2) launched a four-alternative forced choice (4AFC) procedure (Figure 7.5). This was used in the warble discrimination experiment (§9.4).
7.5 Conclusion

Nucleus MATLAB toolbox is a flexible platform for conducting research into cochlear implant sound processing and psychophysics. It was used in the experiments in chapters 8, 9, and 10.

The core modules of NMT (ACE, CIS, psychophysics procedures, and SPrint streaming) have been distributed to several other research groups. Developers of a new strategy can then concentrate on the novel aspects of the processing, without having to re-implement the blocks that are the same as ACE. In Leuven, NMT was used in the development of the F0mod strategy (§6.4.1) (Laneau et al. 2006), and their APEX research software can use the SPrint streaming function (Laneau et al. 2005). In Hannover, NMT and streaming were used in the development and evaluation of the PACE strategy (Nogueira et al. 2005).
8 Place pitch

8.1 Introduction

The percept associated with intra-cochlear electrode position is commonly referred to as place-pitch (§3.6). This chapter describes three experiments exploring place-pitch perception in a group of six Nucleus 24 recipients. The first experiment was the pitch ranking of stimuli on adjacent electrodes (§8.2), to characterise the place-pitch perception of each subject.

The second experiment (§8.3) measured the intermediate pitch of dual-electrode stimuli, following a similar protocol to Kwon and van den Honert (2006). These dual-electrode stimuli were identical to those used by McKay et al. (2003) in their Experiment 2 to validate their loudness model. Although the primary objective of experiment §8.3 was to measure intermediate pitch percepts, the dual-channel stimuli were required to be loudness-balanced in any case, so the opportunity arose to apply the loudness model.

As intermediate place-pitch percepts can be produced by sequential stimulation on two adjacent electrodes, presumably they can also be produced by the normal operation of a sequential stimulation strategy such as CIS or ACE. The goal of the third experiment (§8.4) was to investigate intermediate place-pitch percepts produced by pure tones processed by the ACE strategy.

In §8.5, the centroid model of place-pitch (§3.3) is refined by incorporating the McKay loudness model, and applied to the pure-tone pitch-ranking results (§8.4).

8.2 Pitch ranking of adjacent electrodes

8.2.1 Methods

Each stimulus consisted of an 1800 pps pulse train on a single channel. The duration of each stimulus was 500 ms. The stimulus set encompassed the first seven channels in the recipient's clinical map (i.e. the seven most apical electrodes in use). This set was electrodes E22 to E16 for all subjects except for S05, who used electrodes E21 to E15. For uniformity, the stimuli will be referred to as stimulating channels 1 to 7. Adjacent channels were loudness-balanced then pitch-ranked using the procedures described in §7.4.1 and §7.4.2.

8.2.2 Results

Six Nucleus 24 subjects (S01, S02, S03, S04, S05 and S06 in Table 15.1) participated in this experiment. The results are tabulated in Table 8.1 (with statistics calculated according to §16.3).
and plotted in Figure 8.1 (the abscissa is labelled with the more apical channel of the pair). Subjects S01, S02, S05 and S06 achieved mean scores of 90% or more. For subject S05, the score of 11/16 for channels [2, 3] was significantly lower than his mean score. Subjects S03 and S04 had the lowest scores, being unable to discriminate several pairs of channels. A statistical analysis using the method described in §16.3.2 showed that the spread in scores for subject S04 was significantly greater than would be expected by random fluctuation, with high scores on the channel pairs [4, 5] and [5, 6] contrasting with chance scores on the other channels.

<table>
<thead>
<tr>
<th>Subject</th>
<th>S01</th>
<th>S02</th>
<th>S03</th>
<th>S04</th>
<th>S05</th>
<th>S06</th>
</tr>
</thead>
<tbody>
<tr>
<td>Channels: [1, 2]</td>
<td>8**</td>
<td>8**</td>
<td>12*</td>
<td>7</td>
<td>15**</td>
<td>15**</td>
</tr>
<tr>
<td>[2, 3]</td>
<td>8**</td>
<td>8**</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>14**</td>
</tr>
<tr>
<td>[3, 4]</td>
<td>8**</td>
<td>8**</td>
<td>11</td>
<td>10</td>
<td>15**</td>
<td>15**</td>
</tr>
<tr>
<td>[4, 5]</td>
<td>7*</td>
<td>8**</td>
<td>13*</td>
<td>14**</td>
<td>15**</td>
<td>14**</td>
</tr>
<tr>
<td>[5, 6]</td>
<td>7*</td>
<td>8**</td>
<td>12*</td>
<td>15**</td>
<td>16**</td>
<td>16**</td>
</tr>
<tr>
<td>[6, 7]</td>
<td>7*</td>
<td>8**</td>
<td>14**</td>
<td>9</td>
<td>14**</td>
<td>15**</td>
</tr>
<tr>
<td>Number of trials</td>
<td>8</td>
<td>8</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>Pooled score</td>
<td>94%</td>
<td>100%</td>
<td>74%</td>
<td>68%</td>
<td>90%</td>
<td>93%</td>
</tr>
<tr>
<td>- Significance</td>
<td>&lt;0.005</td>
<td>&lt;0.005</td>
<td>&lt;0.005</td>
<td>&lt;0.005</td>
<td>&lt;0.005</td>
<td>&lt;0.005</td>
</tr>
<tr>
<td>Std dev of scores</td>
<td>7%</td>
<td>0%</td>
<td>11%</td>
<td>19%</td>
<td>11%</td>
<td>5%</td>
</tr>
<tr>
<td>- Pooled estimate</td>
<td>9%</td>
<td>0%</td>
<td>11%</td>
<td>12%</td>
<td>8%</td>
<td>6%</td>
</tr>
<tr>
<td>- Significance</td>
<td>0.55</td>
<td>1.00</td>
<td>0.47</td>
<td>0.02</td>
<td>0.08</td>
<td>0.82</td>
</tr>
</tbody>
</table>

Table 8.1 Pitch ranking results: adjacent channels

Figure 8.1 Pitch ranking results: adjacent channels
8.2.3 Discussion

These results are consistent with previous studies that show substantial variation in electrode discrimination, both between subjects, and across the electrode array within a subject (Tong and Clark 1985; Townshend et al. 1987; Nelson et al. 1995). Those researchers hypothesised that this was due to variations in electrode positioning, current distribution, and neural survival. Cohen et al. (2001) determined electrode location from radiographs for three subjects with a prototype of the Nucleus Contour electrode array, and found that electrode discrimination was enhanced by closeness to the modiolus.

8.3 Pitch ranking of dual-electrode stimuli

8.3.1 Methods

This experiment measured the ability of recipients to perceive intermediate place pitch between two adjacent electrodes. Dual-electrode stimuli were constructed with the same pulse timing as that provided by an 8-maxima ACE map (§7.2.1). The pulse rate on each electrode was 1776 pps. The two electrodes in a stimulus will be referred to as electrode $A$ (the more apical of the pair) and electrode $B$ (the more basal). Viewed as a uniform sequence of 70.4 µs time-slots, the first time-slot contained a pulse on electrode $A$, the second time-slot contained a pulse on electrode $B$, the next six time-slots were empty, and then the sequence repeated. Segments of two such dual-electrode stimuli are shown in Figure 8.2. The standard stimulus configuration was used (monopolar mode, with 25 µs phase width and 8 µs inter-phase gap). The duration of each stimulus was 500 ms.
Initially a pulse train on electrode A alone was loudness-balanced (at a medium-loud level) against a pulse train on electrode B alone. This provided a reference current level for each electrode, denoted $c_a$ and $c_b$. The "mid-point" dual-electrode stimulus was constructed with each electrode initially set to its reference current level. The dual-electrode stimuli on either side of the mid-point had the current level reduced on one electrode to "steer" the pitch towards the other electrode, as shown in Table 8.2. Table 8.2 shows the initial current levels of the stimuli before loudness balancing. To adjust the loudness of a dual-electrode stimulus, the currents on both electrodes were reduced by an equal number of current level steps (thus keeping the ratio of the two currents equal).

<table>
<thead>
<tr>
<th>Stimulus label</th>
<th>Current level electrode A</th>
<th>Current level electrode B</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>$c_a$</td>
<td>-</td>
</tr>
<tr>
<td>-30</td>
<td>$c_a$</td>
<td>$c_b - 30$</td>
</tr>
<tr>
<td>-20</td>
<td>$c_a$</td>
<td>$c_b - 20$</td>
</tr>
<tr>
<td>-15</td>
<td>$c_a$</td>
<td>$c_b - 15$</td>
</tr>
<tr>
<td>-10</td>
<td>$c_a$</td>
<td>$c_b - 10$</td>
</tr>
<tr>
<td>-5</td>
<td>$c_a$</td>
<td>$c_b - 5$</td>
</tr>
<tr>
<td>0 (mid-point)</td>
<td>$c_a$</td>
<td>$c_b$</td>
</tr>
<tr>
<td>5</td>
<td>$c_a - 5$</td>
<td>$c_b$</td>
</tr>
<tr>
<td>10</td>
<td>$c_a - 10$</td>
<td>$c_b$</td>
</tr>
<tr>
<td>15</td>
<td>$c_a - 15$</td>
<td>$c_b$</td>
</tr>
<tr>
<td>20</td>
<td>$c_a - 20$</td>
<td>$c_b$</td>
</tr>
<tr>
<td>30</td>
<td>$c_a - 30$</td>
<td>$c_b$</td>
</tr>
<tr>
<td>B</td>
<td>-</td>
<td>$c_b$</td>
</tr>
</tbody>
</table>

Table 8.2 Initial current levels (before loudness balancing) of dual-electrode stimuli

The experimental procedure used was similar to that of Kwon and van den Honert (2006). Firstly, the two single-electrode stimuli were pitch-ranked (part of experiment §8.2). Next, the mid-point stimulus (labelled "0" in Table 8.2) was loudness-balanced against both single-electrode stimuli (labelled "A" and "B"). Then adjacent stimuli from the set [A, 0, B] were pitch-ranked, i.e. the two pairs of stimuli [A, 0] and [0, B]. An initial block of 8 trials of each pair was done. If the subject scored 7 or 8 out of 8 for a pair, then a new intermediate stimulus from Table 8.2 was introduced to subdivide the interval. The new set was loudness-balanced, and then adjacent stimuli were pitch-ranked. At least 16 pitch-ranking trials were performed for adjacent pairs in the final set.

The three subjects with the highest scores in ranking adjacent single-electrode stimuli (§8.2) were selected to participate in this experiment (i.e. S01, S02, S06 in Table 15.1). To investigate the uniformity of intermediate place pitch perception across the electrode array, three pairs of channels were tested. Each subject was initially tested using the most basal pair of channels in their map (i.e. channels 1 and 2 in §8.2), then retested with channels 6 and 7 (which were also tested in §8.2), and finally with channels 11 and 12. These channels corresponded to identical electrodes for these three subjects (i.e. [E22, E21], [E17, E16], and [E12, E11]).
8.3.2 Loudness balance results

In this section, the McKay loudness model (§3.5.2) will be applied to the dual-electrode loudness balance results. Log loudness will be assumed to be a linear function of current level (Equation 3.6). Because the pulse rate was 1800 pps, the current levels should be well below the knee-point current at which loudness grows more steeply, which typically corresponds to the C-level for a 500 pps pulse train (McKay et al. 2003).

<table>
<thead>
<tr>
<th>Label</th>
<th>( c_1 )</th>
<th>( c_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>208</td>
<td>-</td>
</tr>
<tr>
<td>-20</td>
<td>207</td>
<td>189</td>
</tr>
<tr>
<td>-10</td>
<td>204</td>
<td>196</td>
</tr>
<tr>
<td>0</td>
<td>200</td>
<td>202</td>
</tr>
<tr>
<td>10</td>
<td>194</td>
<td>206</td>
</tr>
<tr>
<td>20</td>
<td>185</td>
<td>207</td>
</tr>
<tr>
<td>B</td>
<td>-</td>
<td>210</td>
</tr>
</tbody>
</table>

Table 8.3 Current levels (after loudness balancing) of dual-electrode stimuli for channels 1 and 2 for subject S01

The model's assumptions will be explained using example data from subject S01. Table 8.3 shows the current levels obtained by subject S01 after loudness balancing the stimuli on channels 1 and 2. A current level of 208 on channel 1 alone (stimulus "A") was the same loudness as current level 210 on channel 2 alone (stimulus "B"). The mid-point stimulus (stimulus "0") was initially set to current levels \((c_1, c_2) = (208, 210)\). The model assumes that the loudness of this mid-point stimulus is the sum of the contributions made by the two individual channels, and that these levels result in each channel making an equal contribution to the loudness; i.e. that each channel is contributing half of the total loudness. A reduction of 8 current levels to \((200, 202)\) was required to match the loudness of the mid-point stimulus "0" to stimulus "A" or "B". The model assumes that both channels still contribute half of the loudness after this reduction. Thus for channel 1, current level 200 produces half of the loudness of current level 208; and for channel 2, current level 202 produces half of the loudness of current level 210. The loudness model for channel 1 is shown in Figure 8.3. Stimulus "A" (current level 208) is arbitrarily assigned a loudness level of 1.0, i.e. a log loudness of zero. Current level 200 therefore has a log loudness of \(\log(0.5) = -0.3\). This pair of loudness-balanced stimuli gives us two points on the log loudness function, i.e. an estimate of the slope \(a\) in Equation 3.6:
Using Equation 3.6, the loudness contribution of each channel in the set of stimuli in Table 8.3 can be calculated. The loudness of each stimulus (which should be a constant) is then simply the sum of the two loudness contributions. In Figure 8.4, the loudness contributions of the two channels are shown as dashed lines, and the sum as a solid line. Note that the single-channel stimuli ("A" and "B") and the mid-point stimulus "0" are constrained to have equal loudness by the construction of the model, so the model's accuracy should be assessed on the four remaining stimuli. The vertical bars shown on those four stimuli indicate the loudness levels predicted by the model for stimuli one current level higher or lower than the subject's loudness-balanced stimulus set in Table 8.3. It can be seen that the model very accurately predicted the balance point for the stimuli "-10" and "+10", and that the error for the stimuli "-20" and "20" was only about one current level.
8.3.3 Pitch ranking results

The pitch ranking percent correct scores were converted to sensitivity index $d'$, and the cumulative $d'$ is shown in Figure 8.5. It is notable that some of the perceived pitch functions are not symmetrical. For example, subject S02 could not distinguish the pitch of the channel [6, 7] "mid-point" stimulus from the pitch of channel 7 alone, despite being able to easily discriminate the mid-point stimulus from channel 6 alone. Subject S06 also had very asymmetric results for channel pairs [6, 7] and [11, 12], with the mid-point stimuli being judged as lower in pitch than the more apical channel of the pair.
8.4 Pitch ranking of pure tones with ACE

In this experiment, subjects pitch-ranked pure tones processed by the ACE strategy. Pure tones were chosen because with quadrature envelope detection they produce steady amplitude patterns, without amplitude modulation (§4.4.3). Thus, only place-pitch cues are available. As the filterbank has overlapping frequency bands, a pure tone can excite several channels. If a large change is made to the frequency of the pure tone, a different set of channels will be stimulated, providing a strong place-pitch cue. For smaller frequency changes, the ratios of the currents on adjacent channels should provide a finer place-pitch cue.

8.4.1 Methods and stimuli

The standard pitch ranking procedure was used (§7.4.2). A set of pure tones was constructed with six-semitone frequency spacing (frequencies of 99, 140, 198, 281, and 397 Hz). Each pure tone was 500 ms in duration, and was turned on and off with 50 ms raised-cosine ramps to minimise transients in the filterbank outputs.

All six subjects who participated in experiment §8.2 (S01, S02, S03, S04, S05, S06 in Table 15.1) also took part in this experiment. The three subjects with the best scores (S01, S02, S05) were subsequently tested with a set of seven tones, spanning the same two-octave range with 4-semitone spacing. One subject (S02) was further tested with a set of 13 tones having 2-semitone
In the pitch-ranking procedure, each subject always compared a pair of stimuli that were adjacent in the set, i.e. all the pairs in one set had a constant frequency ratio (6, 4 or 2 semitones).

The tones were processed by the ACE strategy with 8 maxima and a pulse rate of 1776 pps per channel, implemented in Nucleus MATLAB Toolbox (§7.2.1), and presented to the subject by SPrint streaming (§7.3). The amplitude required for the 250 Hz tone (which is at the centre frequency of the first filter) to produce stimulation that just reached the maximum comfortable current level on the corresponding channel was taken as a reference. All tones were initially set to this reference amplitude. The loudness was subsequently adjusted by adding or subtracting a specified number of current levels to all pulses in a sequence. The lowest frequency stimulus in a set was adjusted to be "medium loud", and then successive adjacent pairs of stimuli were loudness-balanced (§7.4.1). The set was then ready for pitch ranking. A block was initially presented with all stimuli at the loudness-balanced levels. It was subsequently repeated with the loudness roved by randomly adding or subtracting one current level to the second tone in the pair.

Two views of a "sweep" of the 13 tones in the 2-semitone interval stimulus set are shown in Figure 8.6 and Figure 8.7. Each figure shows the compressed envelopes (the LGF output) of the stimuli prior to the set being balanced in loudness. The tone frequencies are labelled on each stimulus. Only the lowest four channels of the map are shown, as no other channel is activated. For pure tones with fundamental frequency of 140 Hz or less, only the first channel (centred at 250 Hz) is activated. This means that after loudness balancing, all stimuli having \( F_0 \leq 140 \text{ Hz} \) should be identical, and therefore chance scores are expected for a pair such as (99, 140) Hz. Nevertheless, such pairs were included in the experiment for two reasons: firstly, they act as a control (if subjects score better than chance on those pairs then something is amiss), and secondly, the same set of fundamental frequencies will be used in later experiments (in chapters 9 and 10) where there are differences between the corresponding stimuli. Figure 8.7 also shows that in the subsequent frequency range, 157 – 250 Hz, only channels 1 and 2 are activated. These are dual-channel stimuli similar to those used in experiment §8.3, so the ratio of the two currents may provide a pitch cue.
8.4.2 Results

The six-semitone results are plotted in Figure 8.8, and the statistics (analysed as described in §16.3) are shown in Table 8.4. Subjects S04 and S06 had chance-level scores for each frequency pair, and for the pooled score. The other four subjects (S01, S02, S03, S05) had pooled scores very significantly above chance. They showed similar non-monotonic patterns in their scores, all four having chance scores for the lowest frequency pair, good scores for the next pair, lower scores for the next pair, and perfect scores for the highest frequency pair. Statistical analysis shows that the standard deviations of their scores (Table 8.4) were significantly larger than would be expected if the scores were independent of frequency.
Table 8.4 Pitch ranking of pure tones: results for 6 semitone intervals.

<table>
<thead>
<tr>
<th>Subject</th>
<th>S01</th>
<th>S02</th>
<th>S03</th>
<th>S04</th>
<th>S05</th>
<th>S06</th>
</tr>
</thead>
<tbody>
<tr>
<td>[99, 140]</td>
<td>8</td>
<td>9</td>
<td>16</td>
<td>15</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>[140, 198]</td>
<td>16**</td>
<td>16**</td>
<td>22*</td>
<td>15</td>
<td>12*</td>
<td>7</td>
</tr>
<tr>
<td>[198, 281]</td>
<td>11</td>
<td>14**</td>
<td>20</td>
<td>23</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>[281, 397]</td>
<td>16**</td>
<td>16**</td>
<td>32**</td>
<td>17</td>
<td>16**</td>
<td>7</td>
</tr>
<tr>
<td>Number of trials</td>
<td>16</td>
<td>16</td>
<td>32</td>
<td>40</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>Pooled score</td>
<td>80%</td>
<td>86%</td>
<td>70%</td>
<td>44%</td>
<td>73%</td>
<td>48%</td>
</tr>
<tr>
<td>- Significance</td>
<td>&lt;0.005</td>
<td>&lt;0.005</td>
<td>&lt;0.005</td>
<td>0.95</td>
<td>&lt;0.005</td>
<td>0.65</td>
</tr>
<tr>
<td>Std dev of scores</td>
<td>25%</td>
<td>21%</td>
<td>21%</td>
<td>9%</td>
<td>21%</td>
<td>14%</td>
</tr>
<tr>
<td>- Pooled estimate</td>
<td>10%</td>
<td>9%</td>
<td>8%</td>
<td>8%</td>
<td>11%</td>
<td>12%</td>
</tr>
<tr>
<td>- Significance</td>
<td>&lt;0.005</td>
<td>&lt;0.005</td>
<td>&lt;0.005</td>
<td>0.22</td>
<td>0.02</td>
<td>0.31</td>
</tr>
</tbody>
</table>

Figure 8.8 Pitch ranking of pure tones: results for 6 semitone intervals.

With four-semitone spacing (Table 8.5 and Figure 8.9), again the scores show a highly significant non-monotonic dependence on frequency, with some similarities amongst the three subjects.
Finally, in the 2-semitone interval result for subject S02 (Table 8.6 and Figure 8.10), again there is substantial variation in scores. To improve the confidence intervals, two additional blocks of trials were performed on the central frequency range 140 – 281 Hz, which had produced two very high scores and the remainder near chance.
Table 8.6 Pure tone pitch ranking results: 2 semitone intervals.

<table>
<thead>
<tr>
<th>Subject</th>
<th>502</th>
</tr>
</thead>
<tbody>
<tr>
<td>[99, 111]</td>
<td>6 / 16</td>
</tr>
<tr>
<td>[111, 125]</td>
<td>8 / 16</td>
</tr>
<tr>
<td>[125, 140]</td>
<td>7 / 16</td>
</tr>
<tr>
<td>[140, 157]</td>
<td>32**/32</td>
</tr>
<tr>
<td>[157, 177]</td>
<td>21 / 32</td>
</tr>
<tr>
<td>[177, 198]</td>
<td>18 / 32</td>
</tr>
<tr>
<td>[198, 223]</td>
<td>19 / 32</td>
</tr>
<tr>
<td>[223, 250]</td>
<td>19 / 32</td>
</tr>
<tr>
<td>[250, 281]</td>
<td>29**/32</td>
</tr>
<tr>
<td>[281, 315]</td>
<td>15**/16</td>
</tr>
<tr>
<td>[315, 354]</td>
<td>11 / 16</td>
</tr>
<tr>
<td>[354, 397]</td>
<td>16**/16</td>
</tr>
</tbody>
</table>

Pooled score 70%
- Significance <0.005

Std dev of scores 22%
- Pooled estimate 11%
- Significance <0.005

Figure 8.10 Pure tone pitch ranking results: 2 semitone intervals

As expected, no subject scored significantly better than chance when both tones in the pair were below 149 Hz. This implies that the experimental procedure, including loudness balancing, was effective. Above 149 Hz, the most striking feature is the non-monotonic dependence of the scores on frequency, which will be analysed in the following section (§8.5). The pattern of results also demonstrates the benefit of the experimental paradigm of ranking pairs with equal intervals across the frequency range. An alternative scheme of choosing one reference frequency, and then varying the comparison frequency, would give widely varying results.
depending on the choice of reference frequency.

Overall, even the best cochlear implant recipient had performance much worse than that of a normally hearing listener. Normally hearing listeners typically have a frequency difference limen of less than 1% for pure tones across this entire frequency range (Moore 1997), and so would be expected to obtain perfect scores with two-semitone spacing (a frequency difference of about 12%).

### 8.5 Applying the centroid model to ACE pure tone results

In this section, the centroid model of place-pitch (§3.6.2) is applied to the pure-tone pitch-ranking results (§8.4) for subjects S01, S02, S03, and S05. Subjects S04 and S06 were omitted because they scored at chance levels, and so fitting a psychometric function would be inappropriate. In the model, sensitivity \( (d') \) is a linear function of the centroid difference (Equation 3.7). Three progressive refinements of the model are presented. Incorporating loudness perception (§3.5) greatly improves the model.

#### 8.5.1 Amplitude centroid

In calculating Equation 3.7, Laneau et al. (2004) used the filterbank envelope amplitudes, but this fails to take into account the mapping from amplitude to stimulus current. In particular, any amplitude values that are below the base-level are discarded and do not produce a stimulation pulse (§4.8). These discarded low amplitudes cannot affect the perceived pitch, so they should be excluded from the centroid calculation.

The top panel of Figure 8.11 shows the amplitude response of the four lowest-frequency filters used in experiment §8.4 (the centre frequencies are 250, 375, 500, and 625 Hz). The remaining filters in the filterbank are not shown because they are not activated for the pure tones below 400 Hz used in experiment §8.4. Amplitude is shown on a linear scale, and frequency is shown on a logarithmic scale (linear in semitones). The response incorporates the effect of base-level, which cuts off the "skirts" of the responses. The bottom panel of Figure 8.11 shows the calculated centroid. The centroid is a relatively smooth function of frequency, except for small steps caused by the base-level truncation of the filter skirts.

The next step in applying the model is to calculate the difference between the centroids for each pair of tones that were pitch-ranked. The frequency interval between pairs of tones was constant within each block of trials. Figure 8.12 shows the centroid difference for the three frequency intervals (6, 4 and 2 semitones). It is plotted in the same style as the pitch-ranking scores: each data point is plotted midway between the frequencies of the two tones.
Figure 8.11 Filter amplitude response and centroid.

Figure 8.12 Amplitude centroid difference
According to the model, the sensitivity should be proportional to the centroid difference (Equation 3.8), so the shape of the centroid difference (as a function of frequency) should be similar to the shape of the scores. The model predicts better scores as frequency increases. Although this describes the general trend of the 6-semitone scores for subjects S03 and S05, and the 4-semitone score for subject S01, it does not explain the non-monotonic behaviour of the scores.

A cumulative normal psychometric function was individually fitted to each subject's pitch-ranking scores using the maximum-likelihood method described by Wichmann and Hill (2001) (§16.5), and scatter plots are shown in Figure 8.13. The fitted psychometric function captures the main trends in the data, but the relatively poor fit is indicated by a wide scatter.

![Figure 8.13 Amplitude centroid model scatter plot](image)

**8.5.2 Compressed-amplitude centroid**

The model can be modified to account for the non-linear compression (§4.8). The top panel of Figure 8.14 shows the response of the first four filters, after compression.
Figure 8.14 Compressed amplitude response and centroid

Figure 8.15 Compressed-amplitude centroid difference
The bottom panel of Figure 8.14 shows the centroid of the compressed amplitudes. Each step corresponds to the activation of an additional channel. The initial steepness of the compression function (Figure 4.13) amplifies the size of the steps. The corresponding centroid differences using compressed amplitudes are shown in Figure 8.15. According to this model, the change in pitch when an additional channel is activated is large compared to the change in pitch as the current varies within a group of channels.

For 6-semitone intervals, the centroid difference has the same shape as the results for subjects S01 and S02. For 4-semitone intervals, the centroid difference has two peaks, similar to the results of subjects S02 and S05 (but not S01), although the model under-estimates the score for the highest frequency pair (315 – 397 Hz). For 2-semitone intervals, the centroid difference provides a very good prediction of subject S02's scores, showing that the peaks in the scores correspond to the steps in the centroid. The fitted psychometric functions and scatter plots for each subject are shown in Figure 8.16. The scatter plot demonstrates the good fit for S02. However, using compressed amplitudes yields a worse fit than uncompressed amplitudes for the 6-semitone scores of subjects S03 and S05, and the 4-semitone score for subject S01.

![Figure 8.16 Compressed amplitude centroid model scatter plot](image)
8.5.3 Loudness centroid

The place-pitch model can be further refined by incorporating the McKay loudness model (Figure 8.17). Firstly an estimate of the log loudness slope was obtained for each subject. For subjects S01 and S02, log loudness slope estimates were obtained from the dual-electrode stimuli of experiment §8.3, which had the same pulse rate (1778 pps) as the ACE pure tone stimuli, and a similar range of current levels. For subjects S03 and S05, who did not participate in experiment §8.3, log loudness slope estimates were obtained from the loudness balancing results of the ACE pure tones themselves. Then using Equation 3.6, the current levels of the stimuli presented to each subject were used to calculate the loudness contribution of each channel, and then the loudness centroid. Finally, the loudness centroid difference was calculated for each pair of stimuli. The corresponding scatter plots are shown in Figure 8.18. The fitted psychometric functions show that the loudness centroid difference was a relatively good predictor of the pitch ranking scores for these four subjects.

![Diagram](image-url)  
**Figure 8.17 Loudness centroid model of place pitch**
8.5.4 Comparison of models

The three models can be compared by examining the deviance (Wichmann and Hill 2001), a numerical goodness-of-fit measure (§16.5). A better fit yields a smaller deviance. The deviances for the three models, for each subject, are shown in Figure 8.19. Although the amplitude model fares well for subjects S01, S03, and S05, it is a poor fit for subject S02. The compressed amplitude model is a very poor fit for subject S03. The loudness model is the most consistent across the four subjects. It gives an excellent prediction of the measured scores for subject S02, who has the most amount of data to be fitted (being the only subject who was tested with two-semitone intervals), as demonstrated in Figure 8.20.
Figure 8.19 Normalised deviance for three place pitch centroid models. Lower deviance indicates a better fit.

Figure 8.20 ACE pure tone results (thin dark blue) and loudness centroid model prediction (thick light blue) for subject S02.
8.5.5 Discussion

Although the loudness centroid model appears useful, it cannot explain all of the results. For example, with dual-electrode stimuli (§8.3), the model predicts that the mid-point stimulus, having approximately equal loudness contributions from both electrodes, should have a pitch mid-way between that of the two single-electrode stimuli. However, 3 of the 9 electrode pairs shown in Figure 8.5 had asymmetrical results. The model may only be applicable in areas of the cochlea where there is uniform neural density and neural-electrode distance.

If the centroid was a smooth, monotonic function of the tone frequency, without the pronounced steps seen in Figure 8.14, then subjects may perceive equal frequency intervals as equal perceptual intervals, which would seem a necessary prerequisite for melody perception. Geurts and Wouters (2004) implemented filters with triangular-shaped frequency responses, and compared them to the standard filters used with the Laura implant, which had flat-topped frequency responses. The triangular filters provided a smoother centroid function. They found better discrimination of the fundamental frequency of synthetic vowels, especially when temporal cues were deliberately suppressed. It may be possible that a smoother centroid could also be provided by a less steep compression function, or by explicitly incorporating the loudness model into the amplitude mapping. Nevertheless, the total available place-pitch range is determined by the positions of the electrodes, and even if the centroid was a linear function of frequency (a straight line in Figure 8.14), it would still have the same end-points. Therefore, smoothing out the steps in the centroid may have the beneficial effect of levelling off the peaks and valleys in pitch ranking measurements such as those in Figure 8.8, Figure 8.9 and Figure 8.10, but it is unlikely to provide any improvement in the average pitch ranking score across the entire frequency range.

It seems likely that the centroid model would apply even if the individual pulses employed current steering (§6.5.2). For example, imagine two successive current-steered pulses with equal amplitude. If the current in the first pulse was steered midway between electrodes 6 and 7, and the current in the second pulse was steered midway between electrodes 7 and 8, then the centroid would be at electrode 7. Thus in the practical situation of sequential stimulation across multiple channels, current steering does not appear to offer any finer control over place pitch.

However, it must also be noted that when intermediate place-pitch cues are produced by overlapping filter bands, they are not very robust. Suppose that we have a low-amplitude tone whose frequency is such that one filter envelope is twice the amplitude of its neighbour, and that this gives rise to an intermediate pitch. If the amplitude of the tone is increased, the larger amplitude envelope will grow more slowly than the smaller amplitude envelope, because of the non-linear compression (§4.8). The ratio of the currents on the two corresponding electrodes will decrease, and the apparent pitch will change.
8.6 Conclusion

Experiment §8.4 demonstrated that intermediate place-pitch percepts can be produced by the normal operation of the ACE strategy, at least for pure tone inputs. The perception of place-pitch cues was predicted quite well by the loudness centroid model developed in §8.5. All three experiments in this chapter confirmed that cochlear implant recipients differ markedly in their ability to use place pitch, and that place pitch perception is often not uniform across the electrode array. The limiting factor on place-pitch perception is likely to be uneven neural survival.

Only low frequency pure tones (below 400 Hz) were used in experiment §8.4, but pure tones throughout the entire frequency range of the processor should be capable of producing intermediate place-pitch percepts. For example, channels 6 and 7, used in the dual electrode experiment (§8.3), will be activated together by a 900 Hz tone, and channels 11 and 12 will be activated together by an 1800 Hz tone. The experiment in chapter 13 will use pure tones in the range 500 – 1000 Hz to address the research question of whether place pitch can convey melody.

It is not known whether recipients can make use of intermediate place-pitch percepts occurring concurrently at multiple places on the electrode array. For example, a harmonic tone with fundamental frequency of 900 Hz will activate adjacent groups of electrodes at each resolved harmonic. However, as demonstrated in §4.6, most tonal sounds produce both place and temporal cues. Pitch perception under these conditions will be investigated in chapter 10.
9 Temporal pitch

9.1 Introduction

This chapter describes three experiments exploring temporal pitch perception in one group of Nucleus 24 recipients. To investigate temporal cues without concurrent place cues, these experiments used stimuli that were presented on a single channel (the most apical channel), with varying pulse timing patterns and pulse rates. Pitch percepts produced by low-rate pulse trains were compared to those produced by modulating a high-rate pulse train. The motivation was to compare the two approaches to enhancing temporal pitch cues described in §6.1, targeting either pulse rate cues or modulation cues. The experiments also addressed the research question of whether the pitch of an amplitude-modulated high-rate pulse train depended only on the fundamental period, or whether the shape of the modulating waveform was important. The stimulation timing resolution required for good temporal pitch cues was investigated. Two models of temporal pitch perception were assessed for consistency with the experimental results.

9.2 Pitch ranking of single-channel modulated sequences

9.2.1 Methods and stimuli

The stimuli were sequences of pulses on the most apical channel. Four types of sequences were used, denoted SPP, MPPU, MPPS, and MPPH, as shown in Figure 9.1. Each panel illustrates one type of sequence. Each row within a panel shows one sequence, with F0 shown on the left axis. The terminology is based on that introduced by Tong, Blamey et al. (1983a) (compare to Figure 3.5).

The Single Pulse per Period (SPP) sequence (Figure 9.1a) was a pulse train with the pulse rate equal to the fundamental frequency, as used in many previous studies of temporal pitch.

The Multiple Pulse per Period, Uniform-sampling (MPPU) sequence was an on-off modulated pulse train (Figure 9.1b). There was one burst of carrier pulses per period of the fundamental frequency. The sequence was generated by deleting selected pulses from a constant rate (1776 pps) carrier sequence (the motivation for this will become clearer in §10.2). Because the fundamental frequency was generally not a sub-multiple of the carrier rate, the number of pulses within a burst, and the intervals between bursts, could vary within a sequence. For example, Figure 9.1b shows that in the 334 Hz MPPU sequence, some bursts contained three pulses, and others contained two. This variation in burst length repeated periodically, with a low frequency (compared to the fundamental frequency), creating a beating effect. In initial testing with these
sequences, some subjects reported that they perceived a "warble". To test the hypothesis that the perceived warble was due to the beating effect, a third set of sequences was constructed, denoted Multiple Pulse per Period, Synchronized (MPPS) sequences (Figure 9.1c). In a MPPS sequence, the start of each burst was synchronized to the fundamental frequency. The time axis was divided into time slots of 70.4 µs, and each pulse was placed in the nearest time slot. The pulse rate within each burst was 1776 Hz (i.e. every 8th time slot within a burst contained a pulse). The number of pulses in each burst, and the interval between bursts, were constant within a MPPS sequence. For example, in the 334 Hz MPPS sequence, all bursts contained three pulses (Figure 9.1c). The MPPS sequences were equivalent to the MPP sequences used in the earlier Melbourne studies described in §3.7 (Tong et al. 1983a; Busby and Clark 1997).

The final sequence type was the Multiple Pulse per Period, Half-wave (MPPH) sequence. An MPPH sequence had exactly the same pulse timing as a MPPS sequence, but the amplitudes of the pulses had a half-wave shape. The motivation was to determine whether the modulation waveform shape affected the perceived pitch.

The standard pitch-ranking procedure was used (§7.4.2). Each block of trials contained sequences of only one type. All stimuli were balanced in loudness before pitch ranking. The frequency interval (in semitones) between the two stimuli in a pair was constant in size in each block of trials.
Figure 9.1 Single-channel pulse sequences.

(a) Single Pulse per Period (SPP);
(b) Multiple Pulse per Period, Uniform-sampling (MPPU);
(c) Multiple Pulse per Period, Synchronized (MPPS);
(d) Multiple Pulse per Period, Half-wave (MPPH).
9.2.2 Results

Seven Nucleus 24 recipients participated in this experiment. They included the six subjects (S01, S02, S03, S04, S05, S06) who took part in experiments §8.2 and §8.4, and one new subject, S07 (Table 15.1).

So that comparisons could be made between place pitch perception and temporal pitch perception (to be described in chapter 10), the same sets of fundamental frequencies were used in this experiment as were used in pitch ranking pure tones processed by ACE (§8.4). The three subjects (S01, S02 and S05) who were tested with 4-semitone intervals for ACE pure tones were also tested with 4-semitone intervals in this experiment. The remaining subjects (and S02 in addition) were tested with 6-semitone intervals. MPPH was not tested with subjects S01 and S04.

The pitch ranking results for 6-semitone intervals are shown in Figure 9.2, and for 4-semitone intervals in Figure 9.3. Overall, the results are consistent with many earlier studies showing that temporal pitch perception degrades as frequency increases beyond 300 Hz (Simmons et al. 1965; Eddington et al. 1978; Tong and Clark 1985; Zeng 2002). As in previous studies, there was substantial variation among subjects.

The upper frequency limit of temporal pitch was not measured, but appeared to be highest for subjects S05 and S06. Subjects S03 and S04 (and S06 to a lesser extent) had reduced scores at low frequencies. This low-frequency degradation has also been observed in temporal modulation detection (Shannon 1992).

The two best-performing subjects (S01 and S06) were also tested with one-semitone intervals. The results are shown in Figure 9.4. As expected, the overall trend is for scores to fall-off as frequency increases. Subject S06 showed a trend for scores to also fall off at low frequencies (consistent with his six-semitone results). The one-semitone plots show more variation between neighbouring pairs of frequencies than was seen with the larger intervals. Scores for subject S01 for MPPS exhibited a very strong pitch reversal for the pair 281 – 297 Hz.

Subject S01 reported anecdotally that the MPPU sequences in the frequency range 200 – 400 Hz had a warble that made the pitch-ranking task difficult. She commented that the MPPS sequences did not have a perceptible warble, and that the pitch-ranking task was easier with MPPS than with MPPU. Subsequently, the other subjects were asked if they could hear a warble in any of the stimuli. Subjects S06 and S07 also reported a warble for some MPPU sequences, but not for the other types of sequences. The remaining subjects could not hear a warble in any of the sequences.
Figure 9.2 Single-channel modulation pitch ranking results: 6 semitone intervals
Figure 9.3 Single-channel modulation pitch ranking results: 4 semitone intervals

Figure 9.4 Single-channel modulation pitch ranking results: 1 semitone intervals
To visualise the effect of modulation type, the group mean results for 6 and 4 semitones are shown in Figure 9.5 (averaged across the subjects who were tested with all four modulation types). Bearing in mind that MPPU and MPPS stimuli are identical except for timing quantisation, it is not surprising that they produced almost identical group mean scores. The SPP and MPPH mean scores were very similar. The MPPU and MPPS mean scores appear to roll off at a lower frequency than the SPP and MPPH mean scores.

Figure 9.5 Single-channel modulation pitch ranking group mean results
top: 6 semitones; bottom: 4 semitones
Temporal pitch

Statistical analysis (as per §16.4.3) of each subject's results, combined across interval sizes, is shown in Table 9.1. SPP produced significantly higher scores than MPPU in four subjects (S01, S02, S06, and S07). No subject showed a significant difference between SPP and MPPH. The only subject who showed a significant difference between MPPU and MPPS was subject S01.

<table>
<thead>
<tr>
<th>Subject</th>
<th>SPP – MPPU</th>
<th>SPP – MPPH</th>
<th>MPPS – MPPU</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean difference (%)</td>
<td>Significance (p-value)</td>
<td>Mean difference (%)</td>
</tr>
<tr>
<td>S01</td>
<td>10</td>
<td>8e-5**</td>
<td>-</td>
</tr>
<tr>
<td>S02</td>
<td>8</td>
<td>7e-3**</td>
<td>-3</td>
</tr>
<tr>
<td>S03</td>
<td>5</td>
<td>0.22</td>
<td>6</td>
</tr>
<tr>
<td>S04</td>
<td>-10</td>
<td>0.98</td>
<td>-</td>
</tr>
<tr>
<td>S05</td>
<td>2</td>
<td>0.37</td>
<td>-1</td>
</tr>
<tr>
<td>S06</td>
<td>6</td>
<td>4e-3**</td>
<td>-3</td>
</tr>
<tr>
<td>S07</td>
<td>20</td>
<td>3e-4**</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 9.1 Paired comparisons of pitch ranking scores for modulation type.

9.2.3 Discussion

The main finding was that pitch ranking scores depended on the sequence type. The on-off modulated high-rate sequences (MPPU and MPPS) produced lower mean scores than the low-rate pulse sequences (SPP), and the pitch percept for MPPU and MPPS sequences appeared to fall off at lower frequencies than for SPP sequences. In contrast, the results for the half-wave shaped modulated high-rate pulse trains (MPPH sequences) were very similar to SPP sequences. Thus, despite having identical pulse timing, the MPPS and MPPH sequences produced significantly different pitch ranking scores.

The results also showed the importance of precise pulse timing. Although not formally investigated in this experiment, it did appear that only the MPPU sequences ever evoked a warble percept. This observation has implications for the pulse timing resolution that is required in a cochlear implant system to convey temporal cues well. Both MPPU and MPPS sequences were intended to provide a temporal cue, but they differed in how accurately they positioned a pulse in time. In the MPPU sequences, the pulse timing was quantized into 563 µs time slots, which appears to be too coarse for some implant recipients. In the MPPS sequences, the pulse timing was quantized into 70 µs time slots, which appears to be acceptable for this group of subjects. An experiment designed to further investigate this issue will be described in §9.4.

Another pulse timing issue was that for the 281 – 297 Hz pair, subject S01 perceived a pitch reversal with the MPPS sequences, but not with the MPPU sequences. To understand this pitch reversal, the MPPS pulse sequences (Figure 9.1c) were closely examined. As the fundamental frequency increased from 281 Hz to 297 Hz, the number of pulses in each burst decremented from four to three. Thus, although the fundamental period decreased, the interval between the bursts (the inter-burst gap) increased. Figure 9.6 shows the MPPS pitch ranking sensitivity $d'$ for
subject S01. It should be a function of the difference in the perceived pitch of the two stimuli. On the same plot is shown the difference (in Hertz) between the fundamental frequencies for the corresponding pairs of stimuli. The fundamental frequency increased monotonically across the stimulus set, so the differences were all positive, but the differences were not monotonically increasing, due to pulse timing quantization. Also shown in Figure 9.6 is the difference between the reciprocals of the inter-burst gaps of each pair of stimuli. The difference was negative whenever the number of pulses in the burst decremented. The largest negative values were associated with dips in the pitch-ranking score, and the most negative value occurred for the pair 281 – 297 Hz, which had the pitch reversal. Clearly, the inter-burst gap had a strong effect on the perceived pitch for this subject.

Figure 9.6 Pitch ranking result for subject S01 with MPPS sequences. Sensitivity ($d'$) is shown on the left ordinate, and frequency differences (in Hz) based on fundamental period and burst gap is shown on the right ordinate. The zero point of the frequency difference axis is aligned with the zero point on the $d'$ axis to represent the expectation of a chance score if there is no frequency difference, and a negative $d'$ (i.e. percent correct score less than 50%) if the difference is negative (a pitch reversal).

### 9.3 Pitch matching of SPP and MPPS sequences

When subjects were asked to describe the sounds in experiment §9.2, they sometimes reported
that a MPPS sequence had a higher pitch than the SPP sequence with the same fundamental frequency. The purpose of the next experiment was to quantify this pitch difference.

9.3.1 Methods and stimuli

A pitch-ranking procedure was used, similar to that of Carlyon, van Wieringen et al (2002) (described in §3.7.1). In each trial, the reference stimulus was the 178 Hz MPPS sequence and the comparison stimulus was an SPP sequence, with pulse rate in the range 125 to 400 pps. The reference MPPS frequency of 178 Hz was chosen because most subjects had relatively good pitch ranking performance at that frequency in experiment §9.2. After several blocks of trials, the set of SPP frequencies was manually adjusted for each subject to straddle the apparent 50% point of the psychometric function.

9.3.2 Results

All seven subjects who participated in experiment §9.2 also took part in this experiment. The results are plotted in Figure 9.7. The abscissa is the pulse rate of the SPP sequence. The ordinate is the percentage of trials that the SPP sequence was ranked higher in pitch than the 178 Hz MPPS sequence. The vertical dashed line indicates the fundamental frequency of the MPPS sequence (178 Hz). Error bars show 90% confidence intervals. A cumulative normal psychometric function was fitted to each subject's results using the maximum-likelihood method described by Wichmann and Hill (2001) (§16.5). Subjects differed markedly in both the slope and 50% point of the psychometric function. Subject S01, who performed best in pitch-ranking the SPP stimuli (§9.2), had the steepest psychometric function. Subject S07 had difficulty with the task and gave less consistent judgements than the other subjects.

The SPP sequence that would best match the pitch of the 178 Hz MPPS sequence was found by taking the 50% point on the psychometric function. Figure 9.8 shows that for each subject, the pitch-match SPP sequence had a period (i.e. inter-pulse interval) that was between the period and the inter-burst gap of the MPPS sequence (indicated by the dashed horizontal lines). The group mean and standard deviation is shown in red. The ability of temporal pitch models to explain this result will be examined in §9.5.
Figure 9.7 Psychometric functions for the pitch of 178 Hz MPPS sequence compared to SPP sequences with differing pulse rates.
9.3.3 Discussion

These results confirmed the anecdotal reports in §9.2 that an MPPS sequence generally had a higher pitch than an SPP sequence with the same period. Because the MPPS pitch-ranking scores generally rolled off at a lower frequency than SPP scores, it suggests that the highest perceptible pitch was about the same for both types of sequence.

These results are consistent with the results of Busby et al. (1997), who found that numerical pitch estimates for MPPS sequences (with frequencies in the range 71 to 250 Hz) were either significantly higher than or equal to those for corresponding SPP sequences in six Nucleus 22 recipients (their Figure 2). In both the Busby study and the present study, there were no cases where a MPPS sequence had a lower pitch than the SPP sequence with equal period.

9.4 Warble discrimination

In experiment §9.2, some subjects heard a warble in MPPU sequences, but no subjects heard a warble in MPPS sequences. The experiment described in this section was designed to measure the sensitivity of recipients to pulse timing resolution.

9.4.1 Methods and stimuli

The subject's task in this experiment was to discriminate between two single-channel sequences that had the same average pulse rate, but differed in pulse timing.

In the first condition, both sequences were multiple-pulse-per-period sequences. The experimental software firstly created a carrier sequence with 70.4 µs time slots, with every 8th
time slot containing a pulse, giving a pulse rate of 1776 pps. The duration was 500 ms. Then the reference stimulus (denoted stimulus R) was constructed by partitioning the carrier pulses into groups of 10, and deleting the last 5 pulses in each group. The beginning of the reference stimulus is plotted as the upper sequence in Figure 9.9. It was a MPPS sequence with a fundamental frequency of 178 Hz (the same frequency used in the MPPS – SPP pitch matching experiment, §9.3). The gap between bursts was 3.31 ms (6 * 8 – 1 = 47 time slots). The variant stimulus, denoted stimulus V, was constructed from the reference sequence by advancing every second burst of pulses by a specified number of time slots. The lower sequence in Figure 9.9 shows the variant stimulus for an advance of 10 time slots (704 µs). In this case, the duration of the gap between bursts alternated between a short value of 2.60 ms (37 time slots) and a longer value of 4.00 ms (57 time slots). Note that both the reference and variant sequences had the same duration (500 ms) and the same number of pulses.

In the second condition, the reference stimulus R was a single pulse-per-period (SPP) sequence with a pulse rate of 178 Hz (upper sequence in Figure 9.10), i.e. a period of 5.63 ms (80 timeslots). The duration was again 500 ms. The variant stimulus V was constructed by advancing every second pulse of stimulus R by a specified number of time slots. For example, the lower sequence in Figure 9.10 shows stimulus V for a time advance of 10 time slots. In this case, the period alternated between 5.56 ms (70 time slots) and 6.34 ms (90 time slots). The sequence was similar in concept to the "4 – 6" sequence used by Carlyon, van Wieringen et al. (2002) (§3.7.1).

Figure 9.9 Warble discrimination stimuli: MPP sequences.
The subject performed a four alternative, forced choice (4AFC) task (§7.4.3). In each trial, there were three presentations of the reference stimulus R, and one presentation of the variant stimulus V, in randomised order (e.g. RRVR). The subject clicked an on-screen button to indicate which of the four stimuli was different. The silent interval between stimuli was 250 ms. No feedback was provided.

Performance was measured as a function of the time advance. Each block of trials used a single value of time advance. The time advance was initially chosen to be sufficiently large to allow high discrimination, and was manually adjusted by the researcher after each block until chance scores were obtained.

### 9.4.2 Results

Four Nucleus 24 recipients participated in this experiment: S01, S06, S07, and S10 (Table 15.1). Subjects S01, S06, and S07 had participated in experiment §9.3, which had involved loudness matching the MPP and SPP sequences, so were tested at the current levels that had been obtained previously (considered medium loud). The new subject S10 adjusted the current levels of the MPP and SPP stimuli separately to be medium loud, but did not perform a formal balancing procedure. Subjects S01 and S06 were also tested at a softer level.

The results are plotted in Figure 9.11. Chance performance (25%) is indicated by the horizontal dotted line; error bars indicate 90% confidence intervals. A cumulative normal psychometric function was fitted to each subject's results (§16.5), and the estimated thresholds are shown in Figure 9.12. Discrimination was worse at the softer level for the two subjects who tested that condition. In most cases, the detection threshold was lower for the MPP sequence than the SPP sequence. The exception was subject S10; a possible explanation is that the loudness of the MPP and SPP sequences may not have been equal. Performance varied widely across subjects. For medium loud SPP sequences, the detection threshold ranged from 260 $\mu$s for S01 to 930 $\mu$s for S07. For medium loud MPP sequences, it ranged from 120 $\mu$s for S06 to 740 $\mu$s for S07.
Figure 9.11 Warble discrimination results for 178 Hz SPP and MPP stimuli. The abscissa is the time advance of alternate SPP pulses or MPP bursts.
Better warble detection at higher current levels is consistent with other studies that have shown better temporal modulation detection at higher current levels (Galvin and Fu 2005; Fu 2002).

These results are relevant to sound processing strategies that seek to provide enhanced temporal cues (Table 6.2). The SPP results are applicable to strategies that use pulse timing cues, such as Peak Derived Timing (§ 6.3.2) or Multirate (§ 6.3.3). Fearn (2001, chapter 4) also investigated this issue. He measured the ability of five recipients to discriminate between a SPP sequence (with inter-pulse period T) and a jittered sequence (with a uniform random distribution of inter-pulse intervals in the range T ± J). For period T = 10 ms, the group mean discrimination threshold was J = 2.9 ms, and the best subject's result was J = 0.8 ms. He concluded that the pulse arbitration scheme could delay a pulse by up to 700 μs before it would be perceptible. However, the pulse arbitration scheme is not the only source of timing deviations. In a strategy that attempts to emit pulses on temporal peaks or on zero-crossings, an inadequate sampling rate of the filter output signal will create timing deviations that occur in a regular pattern. The present results show that (not surprisingly) regular timing deviations are easier to perceive than random deviations.

The MPP results are applicable to strategies that provide an amplitude modulation cue by on-off modulation of continuous high-rate carrier pulse trains, such as HWG (§ 7.2.4) and F0Sync (§ 6.4.2). In this case no pulse arbitration scheme is required, but the size of the timing deviations is related to the stimulation rate.
9.5 Applying temporal pitch models

This section assesses the two different temporal pitch models discussed in §2.4.5 for consistency with the experimental results. To apply these models to cochlear implants, it is necessary to know the neural response to a sequence of electrical stimulation pulses.

The neural response to the SPP and MPPS sequences was simulated following the method of Bruce et al. (1999a; 1999b). Because the precise values of the model parameters are not known, only qualitative descriptions can be given. The loss of inner hair cells results in negligible levels of spontaneous activity in the auditory nerve, so that nerves only fire in response to stimulation pulses. The probability of a nerve firing in response to the \( n \)th pulse within a train of pulses is modelled by:

\[
p(n) = \Phi\left(\frac{I_{stim} - r(n)I_{thresh}}{\sigma}\right)
\]

where \( I_{stim} \) is the stimulus current, \( I_{thresh} \) is the threshold current for a single isolated pulse (at which the firing probability is 50%), \( r(n) \) is a refractory function, \( \sigma \) is the standard deviation of the threshold noise, and \( \Phi \) is the cumulative normal distribution. The refractory function \( r(n) \) represents a threshold shift, and depends on the time since the last spike. It is infinite during the absolute refractory time, and falls exponentially to one during the relative refractory period.

For a SPP sequence, all inter-spike intervals are multiples of the pulse period, and both the autocorrelation and the first-order interval model predict a pitch equal to the pulse rate.

For a MPPS sequence, if the inter-burst gap is long enough for the nerves to fully recover, then the compound neural response to each burst of pulses will be similar. For the 178 Hz MPPS sequence used in experiment §9.3, the inter-burst gap was 3.38 ms. Measurements in Nucleus 24 cochlear implant recipients using neural response telemetry (NRT) indicate that the electrically-evoked compound action potential has typically recovered to 80% of its amplitude after this time (Abbas et al. 1999), so the first pulse in each burst will have a relatively high probability of causing a spike. The simulated response to the MPPS sequence is shown in Figure 9.13. Figure 9.13b shows the individual responses of 20 simulated fibres, with the spikes for each fibre shown as a horizontal row of dots. The compound response for a population of 1000 fibres is shown in Figure 9.13c. Following Cariani and Delgutte (1996), the autocorrelation of each individual fibre response was calculated, and then summed across the population of fibres (Figure 9.14). The summed autocorrelation has a peak at the fundamental period (5.63 ms), and therefore this model predicts a pitch equal to the fundamental frequency. The same result is obtained if the autocorrelation of the compound response is taken. This is not consistent with the results of experiment §9.3. For the MPPS sequence, the pulse rate within the
burst of 1776 pulses per second (inter-pulse interval of 563 µs) is too high for any single neuron to fire on every pulse (Figure 9.14b), and therefore the first-order interval statistics will depend on refractory effects. The distribution of first-order intervals will be broader than the SPP case, and contain shorter intervals. For example, a nerve may fire on only the first pulse in two successive bursts, giving an inter-spike interval equal to the fundamental period. Alternatively, a nerve may fire on the last pulse in one burst and the first pulse in the next burst, giving an inter-spike interval equal to the inter-burst gap. Thus the first-order interval model is consistent with the results of experiment §9.3, where all subjects matched the pitch of the MPPS sequence to that of an SPP sequence with an interval between the fundamental period and the inter-burst gap. The across-subject variation may be due to differing neural refractoriness; for example, subject S05 matched to the fundamental period, perhaps implying that his nerve fibres only fired once for each burst.

The first-order interval model is also consistent with the pitch ranking results of subject S01 for the MPPS sequences (Figure 9.6), where the inter-burst gap had a strong effect on the perceived pitch. This cannot be explained by the autocorrelation model.

![Simulated response to 178 Hz MPPS sequence.](image)

(a) Stimulation pulses.
(b) Responses of 20 individual fibres. Each dot represents a nerve firing.
(c) Spike probability, calculated from the response of 1000 fibres.
These experimental results replicated the well known result that temporal pitch perception degrades at frequencies beyond 300 Hz. Of more interest is the perceptual difference between low-rate pulse trains and amplitude modulated high-rate pulse trains. For most subjects, a high-rate pulse train, modulated on and off at frequency F0, had higher pitch than a train of pulses at the rate of F0. It appeared that using on-off modulation reduced the frequency range over which changes in pitch could be conveyed. If amplitude modulation of high-rate pulse trains was to be used to convey pitch, then the shape of the modulating waveform was important: the half-wave shape was better than the rectangular (on-off) shape.

There were large differences between recipients in their sensitivity to regular patterns of timing deviations in pulse trains: the smallest timing deviations detected varied from 120 μs to 740 μs.

Temporal models for pitch in normal hearing propose that pitch depends on auditory nerve inter-spike intervals (§2.4.5). The warbles sometimes perceived in MPPU sequences, the pitch reversal observed with an MPPS sequence, and the higher pitch of MPPS sequences compared to SPP sequences, all suggest that recipients obtained temporal pitch cues from first-order inter-spike intervals. The results are not consistent with the autocorrelation (all-order inter-spike intervals) model.
10 Concurrent place and temporal cues

10.1 Introduction

This chapter addresses the research question of whether cochlear implant recipients can usefully combine place and temporal cues to pitch.

For pure tone stimuli, ACE with quadrature envelope detection provided no temporal cues to pitch (§4.6). The first experiment in this chapter (§10.2) investigated whether changes to the amplitude detection and sampling scheme to provide temporal cues would improve pitch perception for pure tones. Two schemes were selected to exemplify the two approaches to enhancing temporal cues: half-wave gating (HWG, §7.2.4) and temporal peak sampling (TPS, §7.2.6). HWG is an extreme form of modulating a constant-rate pulse train to provide fine timing information. TPS demonstrates the alternative approach of controlling the timing of individual pulses.

The experiments in §8.4, §9.2 and §10.2 were designed so that the effects of place and temporal cues both alone and together could be compared. The HWG and TPS strategies used the same filterbank as ACE and so provided the same place cues as ACE. In addition, they provided a concurrent temporal cue. The temporal cue in HWG was on-off amplitude modulation, identical to that in the MPPU sequences. The temporal cue in TPS was pulse rate, identical to that in the SPP sequences. It is thus possible to compare the results of the experiments to try to separate the effects of place and temporal cues.

As pure tones are rarely heard outside of the laboratory, the second experiment (§10.3) measured pitch perception with more natural harmonic tones, taken from recordings of sung vowels. These harmonic tones give rise to both place and temporal cues to pitch. Pitch ranking performance was compared using the strategies ACE, HWR, TPS, and F0M (§7.2). The strategies covered all four approaches to improving temporal cues that were defined in Chapter 6 (Table 6.2).

10.2 Pitch ranking of pure tones with HWG and TPS

10.2.1 Methods and stimuli

This experiment measured pitch ranking of pure tones processed by the HWG and TPS strategies, for comparison with the previous ACE results (§8.4). The stimuli were processed off-line by strategies implemented in Nucleus MATLAB Toolbox (§7.2), and the resulting pulse sequences were delivered by SPrint streaming (§7.3). The standard pitch ranking procedure was
used (§7.4.2).

Figure 10.1 combines the block diagrams of earlier figures to show the common elements of the ACE, HWG and TPS strategies; the three strategies differ in their envelope detection and sampling methods. Figure 10.2 shows the stimulation sequences produced by the three strategies in response to a 157 Hz pure tone. As shown in Figure 8.6 and Figure 8.7, this tone excites the first two channels. Because the filters in each strategy are identical, the individual pulses in the HWG and TPS responses have the same amplitude as those in ACE, but the number of pulses and their timing differs. Thus HWG and TPS provide the same place cue as ACE, in addition to providing a temporal cue that ACE lacks. In comparison to the single-channel stimuli in experiment §9.2, the SPP stimuli were equivalent to extracting a single channel of the stimulus produced by TPS processing of a pure tone, and similarly, the MPPU stimuli were equivalent to extracting a single channel of the stimulus produced by HWG processing of a pure tone. Figure 10.2 also shows that the HWG and TPS strategies produce in-phase stimulation on channels 1 and 2, because all bands of the FFT filterbank have the same group delay (i.e. it is a zero-phase filterbank).

Figure 10.1 Processing for one channel: ACE, HWG, TPS
10.2.2 Results

The six subjects who were common to both experiments §8.4 and §9.2 (S01, S02, S03, S04, S05, S06 in Table 15.1) participated in this experiment. Subjects S01, S02, and S05 had been tested with four semitone intervals for ACE processing of pure tones (§8.4), so were also tested with four semitone intervals for HWR and TPS processing. Subjects S03, S04, and S06 were tested only with six semitone intervals. Subject S02 was tested at both four and six semitone intervals.

The results are plotted in Figure 10.3 and Figure 10.4. The ACE results from §8.4 are included for comparison. Scores for all three strategies exhibit some large and non-monotonic variations across frequency. There is a clear trend for HWG and TPS scores to be greater than or equal to ACE scores in all subjects except for a few instances at the highest frequencies. This reflects the expected decline in the benefit of temporal cues as the fundamental frequency approaches 300 Hz. Subjects S04 and S06 had similar scores for both ACE and HWG, but had very different TPS scores. Subject S05 experienced a pitch reversal with TPS for the pair (250, 315) Hz.

Statistical analysis according to §16.4.3 is shown in Table 10.1. Because three paired comparisons were made, only $p$-values less than 0.05/3 should be considered significant. HWG
scores were significantly better than ACE scores in all subjects. TPS scores were significantly better than ACE scores in all subjects except S03 and S05. HWG scores were significantly greater than TPS scores for subjects S04 and S05.

The subjects were also asked to describe the sounds qualitatively. All subjects reported that the lowest frequency HWG tones and TPS tones had lower pitches than the lowest frequency ACE tones. Several subjects expressed a preference for the HWG tones, and reported that they found the lowest frequency TPS tones unpleasant (these stimuli consisted of pulse trains with pulse rates of 140 Hz and lower). Subject S01 reported that HWG tones in the frequency range 200 – 400 Hz had a warble, the same as she heard in the single-channel MPPU sequences (§9.2).

Figure 10.3 Pure tone pitch ranking results, ACE v HWG v TPS (6 semitone intervals)
Figure 10.4 Pure tone pitch ranking results, ACE v HWG v TPS (4 semitone intervals)
<table>
<thead>
<tr>
<th>Subj</th>
<th>Frequency pair (Hz)</th>
<th>ACE score</th>
<th>HWG score</th>
<th>TPS score</th>
<th>ACE %</th>
<th>HWG %</th>
<th>TPS %</th>
<th>HWG-ACE %</th>
<th>TPS-ACE %</th>
<th>HWG-TPS %</th>
</tr>
</thead>
<tbody>
<tr>
<td>501</td>
<td>99-125</td>
<td>10/16</td>
<td>16/16</td>
<td>16/16</td>
<td>63</td>
<td>100</td>
<td>100</td>
<td>38</td>
<td>38</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>125-157</td>
<td>9/16</td>
<td>16/16</td>
<td>16/16</td>
<td>56</td>
<td>100</td>
<td>100</td>
<td>44</td>
<td>44</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>157-198</td>
<td>7/16</td>
<td>16/16</td>
<td>16/16</td>
<td>44</td>
<td>100</td>
<td>100</td>
<td>56</td>
<td>56</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>198-250</td>
<td>9/16</td>
<td>16/16</td>
<td>16/16</td>
<td>56</td>
<td>100</td>
<td>100</td>
<td>44</td>
<td>44</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>250-315</td>
<td>13/16</td>
<td>15/16</td>
<td>15/16</td>
<td>.81</td>
<td>81</td>
<td>94</td>
<td>0</td>
<td>13</td>
<td>-13</td>
</tr>
<tr>
<td></td>
<td>315-397</td>
<td>16/16</td>
<td>15/16</td>
<td>16/16</td>
<td>81</td>
<td>94</td>
<td></td>
<td>-19</td>
<td>-6</td>
<td>-13</td>
</tr>
<tr>
<td></td>
<td>mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td></td>
<td></td>
<td></td>
<td>1e-5**</td>
<td>1e-5**</td>
<td>0.96</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>502</td>
<td>99-140</td>
<td>9/16</td>
<td>15/16</td>
<td>24/24</td>
<td>56</td>
<td>94</td>
<td>100</td>
<td>38</td>
<td>44</td>
<td>-6</td>
</tr>
<tr>
<td></td>
<td>140-198</td>
<td>16/16</td>
<td>16/16</td>
<td>24/24</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>198-281</td>
<td>14/16</td>
<td>16/16</td>
<td>16/24</td>
<td>88</td>
<td>100</td>
<td>67</td>
<td>13</td>
<td>-21</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>281-397</td>
<td>16/16</td>
<td>16/16</td>
<td>24/24</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>99-125</td>
<td>25/48</td>
<td>42/48</td>
<td>24/24</td>
<td>52</td>
<td>88</td>
<td>100</td>
<td>35</td>
<td>48</td>
<td>-13</td>
</tr>
<tr>
<td></td>
<td>125-157</td>
<td>41/48</td>
<td>47/48</td>
<td>24/24</td>
<td>83</td>
<td>98</td>
<td>100</td>
<td>13</td>
<td>15</td>
<td>-2</td>
</tr>
<tr>
<td></td>
<td>198-250</td>
<td>23/48</td>
<td>31/48</td>
<td>22/24</td>
<td>48</td>
<td>65</td>
<td>92</td>
<td>17</td>
<td>44</td>
<td>-27</td>
</tr>
<tr>
<td></td>
<td>250-315</td>
<td>48/48</td>
<td>46/48</td>
<td>20/24</td>
<td>100</td>
<td>96</td>
<td>83</td>
<td>-4</td>
<td>-17</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>315-397</td>
<td>48/48</td>
<td>48/48</td>
<td>24/24</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td></td>
<td></td>
<td></td>
<td>1e-5**</td>
<td>1e-5**</td>
<td>0.57</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>503</td>
<td>99-140</td>
<td>16/32</td>
<td>18/32</td>
<td>13/16</td>
<td>50</td>
<td>56</td>
<td>81</td>
<td>6</td>
<td>31</td>
<td>-25</td>
</tr>
<tr>
<td></td>
<td>140-198</td>
<td>22/32</td>
<td>31/32</td>
<td>13/16</td>
<td>69</td>
<td>97</td>
<td>81</td>
<td>28</td>
<td>13</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>198-281</td>
<td>20/32</td>
<td>28/32</td>
<td>11/16</td>
<td>63</td>
<td>88</td>
<td>69</td>
<td>25</td>
<td>6</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>281-397</td>
<td>32/32</td>
<td>32/32</td>
<td>12/16</td>
<td>100</td>
<td>100</td>
<td>75</td>
<td>0</td>
<td>-25</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td></td>
<td></td>
<td></td>
<td>1e-3**</td>
<td>0.19</td>
<td>0.07</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>504</td>
<td>99-140</td>
<td>15/40</td>
<td>39/48</td>
<td>8/16</td>
<td>38</td>
<td>81</td>
<td>50</td>
<td>44</td>
<td>13</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>140-198</td>
<td>15/40</td>
<td>46/48</td>
<td>10/16</td>
<td>38</td>
<td>96</td>
<td>63</td>
<td>58</td>
<td>25</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>198-281</td>
<td>23/40</td>
<td>34/40</td>
<td>13/16</td>
<td>57</td>
<td>85</td>
<td>81</td>
<td>28</td>
<td>24</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>281-397</td>
<td>17/40</td>
<td>38/40</td>
<td>15/16</td>
<td>43</td>
<td>95</td>
<td>94</td>
<td>52</td>
<td>51</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td></td>
<td></td>
<td></td>
<td>1e-5**</td>
<td>4e-5**</td>
<td>6e-4**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>505</td>
<td>99-125</td>
<td>12/24</td>
<td>23/24</td>
<td>14/16</td>
<td>50</td>
<td>96</td>
<td>88</td>
<td>46</td>
<td>38</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>125-157</td>
<td>19/24</td>
<td>21/24</td>
<td>13/16</td>
<td>79</td>
<td>88</td>
<td>81</td>
<td>8</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>157-198</td>
<td>10/24</td>
<td>22/24</td>
<td>16/16</td>
<td>42</td>
<td>92</td>
<td>100</td>
<td>50</td>
<td>58</td>
<td>-8</td>
</tr>
<tr>
<td></td>
<td>198-250</td>
<td>11/24</td>
<td>21/24</td>
<td>10/16</td>
<td>46</td>
<td>88</td>
<td>63</td>
<td>42</td>
<td>17</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>250-315</td>
<td>19/24</td>
<td>14/24</td>
<td>2/16</td>
<td>79</td>
<td>58</td>
<td>13</td>
<td>-21</td>
<td>-67</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td>315-397</td>
<td>21/24</td>
<td>22/24</td>
<td>16/16</td>
<td>88</td>
<td>92</td>
<td>100</td>
<td>4</td>
<td>13</td>
<td>-8</td>
</tr>
<tr>
<td></td>
<td>mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td></td>
<td></td>
<td></td>
<td>1e-5**</td>
<td>0.05</td>
<td>6e-3**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>506</td>
<td>99-140</td>
<td>6/16</td>
<td>15/16</td>
<td>15/16</td>
<td>38</td>
<td>94</td>
<td>94</td>
<td>56</td>
<td>56</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>140-198</td>
<td>7/16</td>
<td>13/16</td>
<td>15/16</td>
<td>44</td>
<td>81</td>
<td>94</td>
<td>38</td>
<td>50</td>
<td>-13</td>
</tr>
<tr>
<td></td>
<td>198-281</td>
<td>11/16</td>
<td>16/16</td>
<td>15/16</td>
<td>69</td>
<td>100</td>
<td>94</td>
<td>31</td>
<td>25</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>281-397</td>
<td>17/16</td>
<td>14/16</td>
<td>9/16</td>
<td>44</td>
<td>88</td>
<td>56</td>
<td>44</td>
<td>13</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td></td>
<td></td>
<td></td>
<td>1e-5**</td>
<td>1e-5**</td>
<td>0.16</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 10.1 Pure tone pitch ranking results
10.2.3 Combining place and temporal cues

This section addresses the issue of whether the subjects combined the place and temporal cues in predictable ways. If place and temporal cues form separate perceptual dimensions, then it is not clear how a subject will respond when both cues are available in a pitch-ranking task. The recipient may attend to one cue, ignoring the other; or the two cues could interfere with each other; or the two cues could be combined beneficially. If the two cues were combined with equal weight into one composite pitch percept, the sensitivity indices could be additive:

\[ d'_{combined} = d'_{place} + d'_{temporal} \]  

(10.1)

Considering that ACE-processed tones provided only place cues, MPPU and SPP single-electrode sequences provided only temporal cues, and HWG- and TPS-processed tones provided both place and temporal cues, then the corresponding additive sensitivity relationships would be:

\[ d'_{HWG} = d'_{ACE} + d'_{MPPU} \]  

(10.2)

\[ d'_{TPS} = d'_{ACE} + d'_{SPP} \]  

(10.3)

Quantitative predictions are hindered by a ceiling effect: in many cases, the sum of two \( d' \) values was greater than the maximum \( d' \) that could be obtained after applying Equation 2.5. Therefore the results will be discussed qualitatively.

For subject S01, the HWG scores were very similar to the MPPU scores, and the TPS scores were very similar to the SPP scores, implying that the place cue was being ignored when both place and temporal cues were provided. Subject S02 showed an additive relation for both HWG and TPS. For subject S03, the HWG scores showed an additive relation, but the TPS scores were close to the SPP scores (implying that the TPS place cue was ignored). For subject S04, ACE scores were at chance levels (\( d'_{ACE} \) close to zero), so we would expect similar scores for HWG and MPPU, or for TPS and SPP, but this was not found. For subject S05, the HWG score was very close to the MPPU score (implying that the HWG place cue was ignored), and there appeared to be a detrimental interaction between the place and temporal cues in TPS, including a pitch reversal.

In summary, recipients varied in the way they responded when both place and temporal cues were available. In some cases it appears that the recipient only attended to one of the cues, ignoring the other. In some cases the two cues interfered with each other. There were few instances of place and temporal cues being combined in an additive fashion.
10.2.4 Limitations of HWG processing

Since the pulse timing on each channel of the HWG-processed pure tones was the same as that in the MPPU sequences, it should be no surprise that a warble could be heard by at least one subject. In arranging the chapters of this thesis, it seemed natural to report the experiments using temporal cues alone (Chapter 9) before the experiments of this section; however pitch ranking of pure tones with HWG was in fact performed first, and the pitch ranking with MPPU and MPPS sequences was subsequently undertaken to investigate this warble percept, as well as allowing the contributions of place and temporal cues to be measured independently.

The lack of warble in MPPS sequences (§9.2) implies that synchronizing the start of the bursts in HWG processing of tones should also remove the warble. Unfortunately, this would complicate the processing because it would disturb the pattern of interleaving pulses across channels, and would require an arbitration scheme similar to that required in TPS. Another potential method to remove the warble is to increase the stimulation rate (which improves the timing resolution); however this is constrained by the capabilities of the implant. In any case, both proposed modifications fail to address the key issue that temporal pitch appears to depend on first-order inter-spike intervals (§9.5). HWG processing was developed under an assumption that the perceived pitch would depend on the fundamental frequency of the pulse sequence, and that the waveform of the modulating function was not important. This assumption was incorrect: MPPS sequences often have a higher pitch than SPP sequences (§9.3); individual subjects had differing responses to HWG-processed tones and TPS-processed tones; and it is likely that a pure tone processed by HWG will have a higher pitch than if it was processed by TPS. In contrast, MPPH sequences had similar perceptual properties to SPP sequences (§9.2), so using the half-sinusoidal wave shape (HWR) may be preferable to the square (on-off) wave shape (HWG).

10.3 Pitch ranking of sung vowels

10.3.1 Methods and stimuli

In contrast to the pure tones used in the earlier pitch-ranking experiments, the next set of experiments used more natural harmonic tones. A set of sung vowel stimuli had been created at Melbourne University specifically for use in cochlear implant pitch perception studies (Looi et al. 2004; Vandali et al. 2005; Sucher and McDermott 2007). Recordings were made of a male and a female vocalist in a sound-treated room. An electronic keyboard provided a pitch reference for each note, and the vocalist sang the vowels /i/ and /a/. The male vocalist covered the range G2 to E4 (98 – 330 Hz), and the female vocalist covered the range C4 to G5 (262 – 784 Hz). The microphone was at a distance of one metre. The sounds were recorded onto digital
audio tape (DAT), and then subsequently edited on a personal computer. A central 560 ms sample of each vowel was excised and linear onset and offset ramps of 30 ms were applied. The amplitude was adjusted to give equal RMS power for each audio sample.

These stimuli were supplied to the present author as a set of "wav" files. They were processed off-line by strategies implemented in Nucleus MATLAB Toolbox (§7.2), and the resulting pulse sequences were delivered by SPrint streaming (§7.3).

Five Nucleus 24 subjects (S01, S02, S04, S06, and S07 in Table 15.1) participated in this experiment. The standard pitch ranking procedure was used (§7.4.2), except that the stimuli were not loudness balanced prior to pitch ranking. To reduce the number of conditions to be tested and to ensure that temporal cues were available, only the male voice /i/ was used. All five subjects were initially tested with the set of stimuli comprising the four notes G2 (98 Hz), C#3 (139 Hz), G3 (196 Hz), and C#4 (277 Hz), which are spaced at intervals of six semitones. In each trial, the subject ranked a pair of stimuli that were adjacent in this set. There were 24 trials in a block (three pairs × eight repetitions). The two subjects with the highest scores (S01 and S06) were subsequently tested with one-semitone intervals.

Four strategies were tested: ACE, HWR, TPS, and F0M. The ACE, HWR, and F0M strategies used a stimulation rate of 1776 pps per channel and selected 8 maxima per scan. The pulse rate on the high-frequency channels in TPS was also limited to 1776 pps.

Figure 10.5 shows the resulting stimulation patterns produced for the four stimuli with 22-channel ACE (the fundamental frequencies are labelled above each stimulus). As previously discussed in §4.6, the filterbank does not resolve the harmonics. Thus examining which electrodes are stimulated provides little information concerning the fundamental frequency. Instead, the stimulated electrodes fall into two groups, corresponding to the F1 formant (at approximately 300 Hz), and the F2 and F3 formants (at approximately 2300 Hz and 2800 Hz). The formant pattern is largely independent of the voice pitch (§2.6).

Despite the four stimuli having equal RMS power, and being equal in loudness to a normal hearing listener, there are significant differences in the resulting channel amplitudes due to the differing alignment of harmonics, formant frequencies, and filter frequencies. The maxima selection process can further accentuate small amplitude differences because a different set of channels may be selected. The outcome is likely to be loudness differences between stimuli for a cochlear implant recipient. However, the stimuli were not loudness-balanced before pitch-ranking (as had been done in the previous experiments in this thesis) to be representative of typical music listening conditions, where successive tones generally have similar audio level. Thus the results may be influenced by loudness differences introduced by the processing strategies.
Figure 10.6 illustrates the ACE electrodogram for four sung vowels (6 semitone intervals).

10.3.2 ACE six-semitone results

The results for the ACE strategy with 6 semitone intervals will be presented and discussed first. The results are plotted in Figure 10.6, and the statistics (analysed as described in §16.3) are shown in Table 10.2. All five subjects obtained perfect scores for the lowest frequency pair (98 – 139 Hz). The subjects can be divided into two groups based on their performance on the other two pairs. The two good performers (S01 and S06) obtained high scores on all pairs. The three poor performers (S02, S04, and S07) did not score above chance on the middle or high frequency pairs. Indeed, S02 and S07 scored significantly worse than chance for the middle frequency pair (139 – 196 Hz), i.e. a perceived reversal in pitch. The last row of Table 10.2 shows that only the three poor performers had a spread in scores greater than would be expected by chance, i.e. the dependence on frequency was statistically significant for those three subjects.
10.3.3 Applying the centroid model of place pitch

In this section, the place-pitch centroid model described in §8.5 will be applied. For each of the four sung vowels, the compressed amplitude was averaged in time on each channel, and then the centroid was calculated. Loudness parameters were not available for all subjects, and (based on the comparison in §8.5) the extra step of applying the loudness model was not expected to substantially alter the results. The number of channels in the subjects' maps differed: S01 and S07 used 18 channels; S04 used 19 channels; S02 and S06 used 22 channels. The centroids for maps having 18, 19, and 22 channels are plotted in Figure 10.7, and the centroid difference for adjacent pairs of stimuli is shown in Figure 10.8. Clearly, the centroid is not a monotonic
function of fundamental frequency. In particular, the centroid of the 196 Hz stimulus is lower than that of the 139 Hz stimulus in all three maps. If a recipient was using place-pitch to rank the stimuli, then they would be expected to reverse the order of the 139 and 196 Hz stimuli, which is precisely what was observed for subjects S02 and S07.

The ACE sung vowel stimuli contained both temporal and place cues to pitch. It appears that the good performers relied on temporal cues and ignored place cues, whilst the poor performers did the opposite.
10.3.4 Smoothed-ACE results

To test the hypothesis that the two good performers (S01 and S06) relied on temporal cues, they were tested with a strategy that deliberately suppressed F0 modulation, leaving only place-pitch cues. Following the technique used by Laneau et al. (2004), the Smoothed-ACE strategy was identical to the ACE strategy used in §10.3.2, except that a 10 Hz fourth-order Butterworth low-pass filter was applied to the filterbank envelopes (before maxima selection).

The pitch ranking results for the same set of sung vowel stimuli are shown in Figure 10.9. The scores obtained by S01 and S06 (good performers) with Smoothed-ACE were similar to the scores obtained by S02 and S07 (poor performers) using ACE, including a pitch reversal for the 139 – 196 Hz pair. Thus it does appear that the better subjects were able to perceive the place cue, and would use it if there was no other cue available, but they relied on the temporal cue when both cues were available.

![Figure 10.9 Sung vowel pitch ranking results with ACE and Smoothed-ACE (6 semitone intervals)](image)

10.3.5 Comparison of ACE, HWR, TPS, and F0M strategies

In addition to ACE, the same five subjects were tested with the HWR, TPS, and F0M strategies, implemented in Nucleus MATLAB Toolbox as described in §7.2. Two variants of each research strategy were tested, as shown in Table 10.3. The HWR and TPS strategies were tested with the FFT vector-sum filterbank and also with the Lyon cochlear model filterbank (§2.4.2). Both variants of the F0M strategy used the FFT power-sum filterbank, with either a single pulse on each channel per period, or else on-off modulation with 50% duty cycle. The third column of Table 10.3 summarises the approach taken by each strategy to improve temporal cues, as defined in Chapter 6; all four quadrants of Table 6.2 are covered.
Three of the subjects failed to complete every condition: S01 omitted F0M-Single; S02 and S04 omitted HWR-FFT. The results are shown in Figure 10.10, one row per subject, with the ACE results from Figure 10.6 included for comparison. Once again it was found that scores varied non-monotonically with frequency, and there were several instances of pitch reversals. All subjects obtained high scores for the lowest frequency pair with every strategy; however, there was little other consistency among subjects.

Table 10.4 shows the results of statistical comparisons between nine pairs of strategies (listed in the first two rows) for each subject individually, according to the Monte Carlo method described in §16.4.3. Two-sided statistical tests were used, since from the plot of results it was clear that not all differences were in the same direction. A more stringent criterion for significance was applied because of the large number of paired comparisons made (Bonferroni correction): "*" indicates $p < 0.05/9$ and "**" indicates $p < 0.01/9$. Larger, non-significant $p$-values are indicated by ";"; a blank indicates that a condition was not tested with that subject.

The first six columns of Table 10.4 compare each research strategy to the ACE base-line. Subject S01 scored highly with all strategies except the two HWR variants: HWR-FFT and HWR-Lyon produced identical scores and were both significantly different to (worse than) ACE, due to the pitch reversal for the middle frequency pair. In contrast, subject S02 exhibited a pitch reversal with ACE, and obtained significantly better scores with the research strategies, with best performance for HWR-Lyon. There was no statistically significant difference between strategies for subject S04, although there were pitch reversals with TPS-Lyon and F0M-Single at the highest frequency pair. Subject S06 scored near ceiling with all strategies. Subject S07

Table 10.3 Strategies used for sung vowel pitch ranking

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Filterbank</th>
<th>Type of temporal cue</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACE</td>
<td>FFT power sum</td>
<td>Within-channel amplitude modulation</td>
</tr>
<tr>
<td>HWR-FFT</td>
<td>FFT vector sum</td>
<td>Within-channel amplitude modulation</td>
</tr>
<tr>
<td>HWR-Lyon</td>
<td>Lyon cochlear model</td>
<td>Within-channel amplitude modulation</td>
</tr>
<tr>
<td>TPS-FFT</td>
<td>FFT vector sum</td>
<td>Within-channel pulse timing</td>
</tr>
<tr>
<td>TPS-Lyon</td>
<td>Lyon cochlear model</td>
<td>Within-channel pulse timing</td>
</tr>
<tr>
<td>F0M-OnOff</td>
<td>FFT power sum</td>
<td>Global amplitude modulation</td>
</tr>
<tr>
<td>F0M-Single</td>
<td>FFT power sum</td>
<td>Global pulse timing</td>
</tr>
</tbody>
</table>

Table 10.4 Sung vowel pitch ranking: significance for paired comparisons of strategies (6 semitone intervals)

<table>
<thead>
<tr>
<th></th>
<th>ACE</th>
<th>ACE</th>
<th>ACE</th>
<th>ACE</th>
<th>ACE</th>
<th>HWR</th>
<th>TPS</th>
<th>F0M</th>
<th>HWR</th>
<th>TPS</th>
<th>F0M</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FFT</td>
<td>Lyon</td>
<td>FFT</td>
<td>Lyon</td>
<td>FFT</td>
<td>On Off</td>
<td>Single</td>
<td>Lyon</td>
<td>Lyon</td>
<td>On Off</td>
<td></td>
</tr>
<tr>
<td>S01</td>
<td>**</td>
<td>**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S02</td>
<td>**</td>
<td>**</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S04</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S06</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
had a pitch reversal with ACE and both HWR variants, and had significantly better scores with TPS-FFT and both F0M variants.

The last three columns of Table 10.4 compare the two variants of each research strategy. There was no significant difference in performance between the FFT filterbank and the Lyon filterbank for either the HWR or TPS strategies. The modulation type used in the F0M strategy appeared to make little difference. The possible exception is subject S02 at the highest frequency, who obtained a perfect score with F0M-Single, contrasting with a chance score for F0M-OnOff. This is consistent with the earlier result using single-channel modulation (§9.2) where subject S02 had the largest decline in scores from SPP to MPP sequences at high frequencies (Figure 9.2).
Figure 10.10 Sung vowel pitch ranking results with ACE, HWR, TPS, and F0M
(6 semitone intervals)
10.3.6 One-semitone intervals

The results of subjects S01 and S06 with one-semitone intervals with ACE, HWR-Lyon, TPS-Lyon, and F0M-OnOff are plotted in Figure 10.11. Paired comparison statistics are listed in Table 10.5 (as per Table 10.4). The stimuli were restricted to the frequency range where temporal cues should be most effective (98 – 233 Hz). Compared to the ACE base-line strategy, subject S01 performed equivalently with F0M-OnOff, but significantly worse with HWR-Lyon and TPS-Lyon. There was no significant difference between strategies for subject S06.

![Figure 10.11 Sung vowel pitch ranking results with ACE, HWR, TPS, and F0M (1 semitone intervals)](image)

<table>
<thead>
<tr>
<th></th>
<th>ACE</th>
<th>ACE</th>
<th>ACE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HWR Lyon</td>
<td>TPS Lyon</td>
<td>F0M On Off</td>
</tr>
<tr>
<td>S01</td>
<td>**</td>
<td>**</td>
<td>-</td>
</tr>
<tr>
<td>S06</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 10.5 Sung vowel pitch ranking: significance for paired comparisons of strategies (1 semitone intervals)

10.3.7 Discussion

With these sung vowel stimuli, place cues were dominated by the formant pattern and were often unrelated to fundamental frequency. Conversely, temporal cues were generally a reliable guide to fundamental frequency. Thus, the best-performing subjects (S01 and S06) were those
who attended to the temporal cues, and ignored the place cues. For these better subjects, the amplitude modulation provided by the ACE strategy was a sufficient temporal cue, and the research strategies did not provide improved performance. The poorer-performing subjects (S02, S04, and S07) had scores with ACE that were at or below chance for the middle frequency pair (139 – 196 Hz). It is encouraging that for each of these subjects, at least one of the research strategies provided scores near ceiling for that frequency pair, indicating that the temporal cue did become more salient. The most consistent strategy in this regard was F0M. Even so, results for the 196 – 277 Hz pair were generally lower than with single channel stimuli at the same frequencies (§9.2), perhaps indicating that subjects still had difficulty in using the temporal cue in the presence of the distracting place cue.

The occurrence of pitch reversals (scores significantly below chance) again demonstrates the prudence of using the method of constant stimuli. Adaptive rules to vary the frequency interval may fail to converge if the underlying psychometric function is not monotonic (§5.2.1).

10.4 Conclusion

These experiments investigated pitch perception when both place and temporal cues were available. For pure tones, subjects obtained better pitch ranking scores with the HWG and TPS strategies (which provided both place and temporal cues), than with ACE (which provided only place cues). Despite this, there were few cases of cochlear implant recipients combining place and temporal cues in an additive fashion. It should be emphasised that pure tones are quite unrealistic sounds, and the main motivation for their use was to allow comparisons to be made between pitch ranking using place cues alone and temporal cues alone.

With the more realistic sung vowel stimuli, the standard ACE strategy provided amplitude modulation cues to pitch for fundamental frequencies in the range 100 – 300 Hz which the best subjects were able to make use of. In contrast, place cues indicated the formants, not the fundamental frequency. The centroid model of place-pitch predicted the pitch reversals observed with ACE in the poorer-performing subjects. The enhanced temporal cues provided by the research strategies improved pitch-ranking scores in these subjects. In contrast, the best-performing subjects were capable of extracting pitch from the temporal cues in ACE, and the research strategies provided no benefit.

Based on their pitch-ranking scores for six-semitone intervals, the three poorer performing subjects appear to have dim prospects for music appreciation with any of the strategies tested. Conversely, the level of performance achieved by the two best subjects for pitch-ranking one-semitone intervals appears sufficient for melody perception. However, there are at least two caveats. Firstly, the stimuli used here are not representative of typical music, which has multiple instruments playing different parts. Secondly, an ability to rank the pitch of two tones does not
necessarily reflect an ability to perceive a melody. The next chapter addresses the issue of measuring melody perception directly.
11 The Modified Melodies test

11.1 Introduction

This chapter describes the Modified Melodies test, which was developed to address some limitations of existing pitch perception tests when used with cochlear implant recipients. One drawback of the forced-choice pitch ranking procedure used in previous chapters is that there is no guarantee that subjects are using melodic pitch in their judgments. Subjects may base their decision on some other aspect of the sound that can be ranked, such as brightness. The objective of this work was to develop a test that could measure pitch perception in accordance with the definition given in §2.2: "that attribute of sensation whose variation is associated with musical melodies" (Plack and Oxenham 2005a).

The new test was firstly trialled with normal hearing subjects. It was then employed with cochlear implant recipients, using their usual sound processor. It was also used in a University of Melbourne Master of Audiology thesis (Stewart 2006) to investigate the benefit to cochlear implant recipients of a contralateral hearing aid, and those results are also reported here with permission.

11.2 Methods

The Modified Melodies test software was written by the present author in the Python programming language. In each trial, the name of a familiar melody was displayed to the subject (as shown in Figure 11.1).

![Figure 11.1. The subject's window of the Modified melodies software](image)

When the subject pressed the Play button, the melody was presented twice. In one of the presentations, randomly selected, the melody was deliberately modified. The subject's task was to select the un-modified (correct) melody by clicking the corresponding button. No feedback
was given. The modification only affected the pitch, and the melodies were presented with rhythmic cues intact. The difficulty of the test was controlled by changing the amount of pitch modification.

11.2.1 Stimuli

The default set of melodies consisted of the initial bars of Old MacDonald and Twinkle Twinkle Little Star. Each melody was defined by a sequence of notes. The pitch of each note was specified as the number of semitones above a base frequency (typically C4, 262 Hz). The melodies were transposed so that their lowest note was equal to the base frequency (i.e. 0). The duration of each note was specified as a number of beats. A beat was 300 ms in duration. Old MacDonald was defined as:

\[
\text{notes} = [5,5,5,0, 2,2,0, 9,9,7,7, 5] \\
\text{beats} = [1,1,1,1, 1,1,2, 1,1,1,1, 2]
\]

Twinkle Twinkle Little Star was defined as:

\[
\text{notes} = [0,0,7,7, 9,9,7, 5,5,4,4, 2,2,0] \\
\text{beats} = [1,1,1,1, 1,1,2, 1,1,1,1, 1,1,2]
\]

Both melodies span a 9-semitone range, and are plotted in Figure 11.2 and Figure 11.3 (as described in §2.7.2). Time (in beats) is indicated on the horizontal axis. The vertical axis shows fundamental frequency on a logarithmic scale (i.e. linear in semitones), with note names indicated.

A later version of the software (used in chapter 13) had the ability to randomly vary the starting note of the melodies across trials. On each trial, the correct and modified versions of the melody were transposed together (i.e. they had the same starting note) by 0, 1, 2 or 3 semitones.
The software allowed a choice of pure or harmonic tones at several frequency ranges. The tones were synthesised at a sampling rate of 16 kHz. Each tone had a smooth (sinusoidal-shaped) rise and fall time of 50 milliseconds. The harmonic tones all had the same spectral profile (i.e. the
same timbre). They were synthesised by generating a train of digital pulses at the fundamental frequency, then applying a band pass filter with lower and upper 3 dB corner frequencies of 200 and 1000 Hz respectively. After filtering, the tones were normalised to have equal amplitude. The filter was designed in MATLAB as a 4th order Butterworth IIR filter, however, only FIR filters were available in the Python "numarray" software library, so it was implemented as an FIR filter with the impulse response truncated to 131 samples. The resulting line spectrum for a 200 Hz harmonic tone is illustrated in Figure 11.4. The first half-dozen harmonics, which dominate pitch perception in normal hearing (§2.4.4), have the highest amplitude. Spectrograms of several examples of these tones were given in §4.6 (Figure 4.11 and Figure 4.12).

![Figure 11.4. Frequency response of band-pass filter (green) and resulting line spectrum of 200 Hz harmonic tone (black)](image)

Four types of pitch modification were used, as described in the following sections. In each case, the rhythm of the melody was unchanged.

**11.2.2 Stretch distortion**

In the Stretch distortion, all the musical intervals in the melody were stretched or compressed. The amount of distortion was specified by a *stretch factor*. The distorted melody was obtained by calculating the intervals (in semitones) between successive notes of the original melody, multiplying these intervals by the stretch factor, then using the new intervals to construct a new melody, starting from the original first note. Stretch factors greater than 1.0 produce stretched intervals, and stretch factors in the range 0 to 1 produce compressed intervals. For example, applying a stretch factor of 2 (doubling all the intervals) to Old MacDonald:

```python
original_notes = [5, 5, 5, 0, 2, 2, 0, 9, 9, 7, 7, 5]
original_intervals = [0, 0, -5, 2, 0, -2, 9, 0, -2, 0, -2]
```
The Modified Melodies test 179

$$\text{stretched_intervals} = [0, 0, -10, 4, 0, -4, 18, 0, -4, 0, -4]$$
$$\text{modified_notes} = [5, 5, 5, -5, -1, -1, -5, 13, 13, 9, 9, 5]$$

The first and last notes are unchanged, and the frequency range has doubled from 9 to 18 semitones. This type of melody distortion was used by Moore and Rosen (1979) in a closed-set melody identification task (§5.2.3).

Figure 11.5 is a plot of Old MacDonald for a stretch factor of 1.11 (shown in red), overlayed on the original melody (shown in black). This stretch factor was used in experiments with normal hearing subjects (§11.2.6). Figure 11.6 is a similar plot of Twinkle Twinkle Little Star. The modified melodies contain "mistuned" notes that lie between the notes of the musical scale, and cannot be represented in standard musical notation. The test becomes more difficult as the stretch factor approaches 1.0, with a chance score expected for a stretch factor of 1.0 (because the “distorted” melody is then identical to the correct melody).

A possible non-melodic cue is the overall frequency range. For example, if the stretch factor was greater than one in every trial, the subject could potentially identify the correct melody merely because it had the smaller frequency range. To avoid this, each block of trials contained both stretched and compressed intervals in reciprocal pairs (e.g. 0.90 and 1.11), in randomised order.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure11_5}
\caption{Old MacDonald Stretch 1.11 distortion (red) overlayed on original (black)}
\end{figure}
11.2.3 Nudge distortion

With the Nudge distortion, a small number of notes in the melody were changed in pitch, and the remaining notes were undisturbed. A defect list defined the notes that were to be changed, and was specified separately for each melody. The defect list was multiplied by a nudge factor, to give an offset list (in semitones), and then the offsets were added to the notes of the original melody. For example, the calculations for a nudge factor of 5 applied to Old MacDonald are:

Original notes = [5, 5, 5, 0, 2, 2, 0, 9, 9, 7, 7, 5]
Defect list = [0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0]
Offset list = [0, 0, 0, 0, 5, 5, 0, 0, 0, 0, 0, 0]
Modified notes = [5, 5, 5, 0, 7, 7, 0, 9, 9, 7, 7, 5]

This gives the modified melody shown in Figure 11.7, with a five semitone shift of the specified notes (DD shifted to GG). The calculations for the same nudge factor of 5 applied to Twinkle Twinkle Little Star are:

Original notes = [0, 0, 7, 7, 9, 9, 7, 5, 5, 4, 4, 2, 2, 0]
Defect list = [0, 0, -1, -1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
Offset list = [0, 0, -5, -5, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
Modified notes = [0, 0, 2, 2, 9, 9, 7, 5, 5, 4, 4, 2, 2, 0]

Note that the defect list for Twinkle Twinkle Little Star was defined with negative signs, so that the notes were shifted down in frequency (in contrast to Old MacDonald), as shown in Figure 11.8. In the experiments described in the following sections, only a single defect list (shown...
above) was used for each melody, i.e. the defect always occurred at the same time position.

Figure 11.7. Old MacDonald Nudge 5 distortion (red) overlayed on original (black)

Figure 11.8. Twinkle Twinkle Nudge 5 distortion (red) overlayed on original (black)
These two melodies and the two corresponding defect lists were selected in an attempt to make the difficulty of the task similar for both melodies. For both melodies:

- Positive nudge factors shift two repeated notes towards the centre of the frequency range.
- Nudge factors of –2, 2, 3, 5, 7 keep the notes in the C major key (i.e. the white keys on a piano).
- A nudge factor of +7 shifts the notes to one extreme of the frequency range of the melody.
- A nudge factor of –2 shifts the notes to the other extreme of the frequency range of the melody.

For nudge factors in the range –2 to +7 the overall frequency range of the melody is unchanged, in contrast to the stretch type of distortion. There is therefore no concern in having a block of trials that contains a single nudge factor.

It was not certain that the difficulty would be a monotonic function of the nudge factor. For comparison, Trainor and Trehub (1992) asked normal hearing adults to detect a change to one note in a melody. The subjects detected a one semitone change (which went outside the key) more easily than a four-semitone change (which remained within the key).

In the case of fractional nudge factors (less than 1.0), the Modified Melodies test clearly does become more difficult as the nudge factor approaches zero, with a chance score expected for a nudge factor of zero (because the “distorted” melody is then identical to the correct melody).

Normal hearing subjects (§11.3.5) were tested with nudge factors 2, 1, 0.5, and 0.25. Cochlear implant recipients (§12.4.2, §13.3) were tested with nudge factors 2, 4, 5, and 6.

### 11.2.4 Exchange distortion

Some cochlear implant recipients were unable to score above chance with the Nudge distortion (§12.4.2), so an "easier" (more distorted) condition was developed, in which a pair of repeated notes was exchanged in order. With the intention of making the difficulty of the task similar for both melodies, the same notes (D and G) were exchanged in the two default melodies, as shown in Figure 11.9 (Old MacDonald) and Figure 11.10 (Twinkle Twinkle Little Star). Looking at it another way, there are two 5-semitone defects in each melody, and the first such defect matches the defect in the corresponding Nudge 5 distortion. This provides a point of comparison between the two distortion types: scores for this Exchange distortion would be expected to be higher than scores for the Nudge 5 distortion.
11.2.5 Backwards distortion

In the Backwards distortion, the notes in the melody were reversed in time. In generating the
modified melody, repeated notes were treated as a single note, as illustrated in Table 11.1. The reversed note sequence was then split into multiple notes to reproduce the rhythm of the original melody. The resulting modified melodies are shown in Figure 11.11 (Old MacDonald) and Figure 11.12 (Twinkle Twinkle Little Star). Because so many notes differ, the modified melodies are plotted in separate panels in this case.

<table>
<thead>
<tr>
<th>Melody</th>
<th>Original note sequence</th>
<th>Backwards note sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Old MacDonald</td>
<td>FCDCAGF</td>
<td>FGACDCF</td>
</tr>
<tr>
<td>Twinkle Twinkle Little Star</td>
<td>CGAGFEDC</td>
<td>CDEFGAGC</td>
</tr>
</tbody>
</table>

Table 11.1 Backward melodies note sequences

Figure 11.11. Old MacDonald (black - bottom) and Backwards distortion (red - top)
11.2.6 Comparison with other test procedures

The lessons learnt from the review of other tests of pitch and melody perception (§5.2) were reflected in the design of the Modified Melodies test.

Because the brightness attribute of timbre can be ranked from low to high (§2.5), it could yield high scores in pitch ranking tasks, and consistent relationships in pitch scaling and estimation tasks. Instead, the Modified Melodies test directly incorporates melody perception.

Using melodies in a discrimination task is unsatisfactory, because there is no way to ensure that the distinguishing stimulus characteristic is actually perceived as pitch. For example, a melody and its modified version may sound different to each other because of some discrepancy in the sound quality of their differing notes (roughness, brightness, or even loudness), and yet neither may convey the correct sense of melody. Furthermore, examples of pitch reversals were seen in pitch ranking sung vowels (§10.3), where a subject consistently ranked the stimulus having higher F0 as lower in pitch. Such pitch reversals have been noted by other researchers (Vandali et al. 2005; Sucher and McDermott 2007). Clearly the subject could discriminate between that pair of stimuli, but the correct pitch was not being conveyed. Instead, the Modified Melodies task asks the subject which presentation of the melody sounds correct. The Modified Melodies test is based on similar principles to the Distorted Tunes Test (§5.2.6), but unlike the Distorted Tunes Test, a forced choice task is used. This avoids subject bias (§2.3.1), allows the test to be
used with a much smaller set of melodies, and provides the means to explore interval and contour aspects of melody perception by varying the pitch modifications.

In common with all tests that use familiar melodies, a drawback of the Modified Melodies test is that it relies on the subject's long-term memory of the melodies. It is intended to be used with subjects who had relatively normal hearing earlier in their life. However, it does not need a large set of familiar melodies, and indeed a single melody is sufficient (as will be shown in §13). This makes it feasible to test cochlear implant recipients who have relatively little musical experience. The task is structured so that rhythm cues are intact (retaining the familiarity of the melodies), but provide no assistance.

Some pitch perception tests are simply too difficult for the typical cochlear implant recipient. It is very disheartening for the subject if they have to guess on every trial. Given the large variation in performance levels between cochlear implant recipients, the ability to adjust the difficulty of the task is extremely useful. Every cochlear implant recipient who was tested with the easiest Modified Melodies condition (the Backwards distortion) was able to score significantly above chance (most of them close to 100%). Conversely, the Modified Melodies test can be made difficult enough to avoid ceiling effects with normal hearing subjects, as will be demonstrated in §11.3.

11.3 Normal hearing subjects

The Modified Melodies test was initially used with normal hearing subjects. The goals were to refine the test procedure, validate the software, and collect normative data, before use with cochlear implant recipients. Fourteen subjects were tested by the author in Sydney. Their details are shown in Table 11.2. In the final column, an "M" indicates those subjects who had spent at least 5 years learning and playing a musical instrument. Subject NH01 was the author. Subjects NH03 and NH04 were the author's children. Five additional subjects (NH12 – NH16, four females and one male) were tested by Penny Stewart in Melbourne. They were selected for their lack of musical experience, and ranged in age from 22 to 24 (Stewart 2006). All subjects had self-reported normal hearing.
11.3.1 Stretch results: 8-melody set

Pilot testing with normal hearing subjects utilised the set of 8 melodies shown in Figure 11.13, which had been rated as the most familiar melodies for Australian listeners (Looi et al. 2003). In contrast to a familiar melody identification task (§5.2.3), the presence of melodies with distinctive rhythms (such as Happy Birthday and Silent Night) and different durations is of no concern. The initial test protocol is summarised in Table 11.3. Each block contained 16 trials (8 melodies × 2 stretch factors). The comparison between performance with harmonic tones and pure tones was done with the most difficult stretch factors (0.95, 1.05) to avoid ceiling effects. Additional blocks of trials were performed if time permitted to improve the confidence intervals.

<table>
<thead>
<tr>
<th>ID</th>
<th>Gender</th>
<th>Age</th>
<th>Musical experience</th>
</tr>
</thead>
<tbody>
<tr>
<td>NH0 1</td>
<td>M</td>
<td>44</td>
<td>-</td>
</tr>
<tr>
<td>NH0 2</td>
<td>F</td>
<td>41</td>
<td>M</td>
</tr>
<tr>
<td>NH0 3</td>
<td>M</td>
<td>12</td>
<td>-</td>
</tr>
<tr>
<td>NH0 4</td>
<td>F</td>
<td>10</td>
<td>-</td>
</tr>
<tr>
<td>NH0 5</td>
<td>F</td>
<td>74</td>
<td>-</td>
</tr>
<tr>
<td>NH0 6</td>
<td>F</td>
<td>24</td>
<td>M</td>
</tr>
<tr>
<td>NH0 7</td>
<td>M</td>
<td>27</td>
<td>M</td>
</tr>
<tr>
<td>NH0 8</td>
<td>M</td>
<td>26</td>
<td>M</td>
</tr>
<tr>
<td>NH0 9</td>
<td>M</td>
<td>38</td>
<td>M</td>
</tr>
<tr>
<td>NH1 10</td>
<td>M</td>
<td>34</td>
<td>M</td>
</tr>
<tr>
<td>NH1 11</td>
<td>F</td>
<td>28</td>
<td>M</td>
</tr>
<tr>
<td>NH1 12</td>
<td>M</td>
<td>23</td>
<td>M</td>
</tr>
<tr>
<td>NH1 13</td>
<td>F</td>
<td>43</td>
<td>-</td>
</tr>
<tr>
<td>NH1 14</td>
<td>F</td>
<td>31</td>
<td>M</td>
</tr>
</tbody>
</table>

Table 11.2 Normal hearing Sydney subject details
Table 11.3 Modified Melodies protocol for normal hearing subjects

<table>
<thead>
<tr>
<th>Block</th>
<th>Tone</th>
<th>Base Freq (Hz)</th>
<th>Distortion</th>
<th>Factors</th>
<th>Num repeats</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Harmonic</td>
<td>262</td>
<td>Stretch</td>
<td>0.75, 1.33</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>Harmonic</td>
<td>262</td>
<td>Stretch</td>
<td>0.90, 1.11</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>Harmonic</td>
<td>262</td>
<td>Stretch</td>
<td>0.95, 1.05</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>Pure</td>
<td>262</td>
<td>Stretch</td>
<td>0.95, 1.05</td>
<td>1</td>
</tr>
</tbody>
</table>
Figure 11.13 Most familiar melodies (in alphabetical order) for Australian listeners.
The results for four subjects are plotted in Figure 11.14. Each panel shows one subject's scores. The symbols show the percent correct scores plotted as a function of stretch factor on a logarithmic axis. Chance performance (50%) is indicated by the dotted horizontal line. The dotted vertical line denotes stretch factor 1.0, for which chance performance is expected. The 90% binomial confidence intervals are indicated by tick marks, so that if the lower tick is above 50% it means that the score was significantly above chance according to a one-sided binomial test (p < 0.05).

The expected pattern of decreasing scores as the stretch factors approached 1.0 was apparent. Each subject was asked to comment on the difficulty of the task, and all reported greater difficulty with pure tones. A statistical comparison of the harmonic and pure tone results is given in §11.3.3.

As the primary intention was to perform within-subject comparisons, it was not strictly necessary for all subjects to use the same set of melodies. It was however important that each subject was familiar with all the melodies in their set, as the intent was for the subject to compare the melodies heard with their long-term memory of the melody. If the melody was unfamiliar, the subject's task would instead be to detect the presence of the out-of-key notes, which was considered too difficult for cochlear implant recipients.

Pilot testing and discussions with cochlear implant recipients indicated that some of these melodies were not sufficiently familiar (see §11.4). It was also apparent that some melodies were more difficult than others. For example, Jingle Bells opens with the same note repeated seven times, so the stretch distortion only affects the last four notes of the excerpt. Subsequently it was decided to reduce the set to the two melodies mentioned in §11.2.1: Old MacDonald and Twinkle Twinkle Little Star.
Figure 11.14 Individual stretch results: 8-melody set, 262 Hz harmonic and pure tones
11.3.2 Stretch results: 2-melody set

A larger study with 15 normal hearing subjects used the default melodies (Old MacDonald and Twinkle Twinkle Little Star), with base frequency of C4 (262 Hz). The test protocol is summarised in Table 11.4. Each block contained 16 trials (2 melodies × 2 stretch factors × 4 repeats). The stretch factors (0.75, 1.33) were omitted, because they had produced perfect scores with the 8-melody set, and were perceived as "too easy". Thus, subjects were initially tested with harmonic tones having the stretch factors (0.90, 1.11). The next block used the stretch factors (0.95, 1.05). If time permitted and the subject was scoring significantly above chance, subsequent blocks were done with stretch factors (0.97, 1.03) and (0.99, 1.01), stopping when chance scores were obtained. For most subjects, additional blocks of trials were then performed at the previous stretch factors to improve the confidence intervals of the results. Harmonic tones were tested with all 15 subjects. For a subset of 9 subjects, pure tones were also tested for at least the pair of stretch factors (0.95, 1.05).

<table>
<thead>
<tr>
<th>Block</th>
<th>Tone</th>
<th>Base Freq (Hz)</th>
<th>Distortion</th>
<th>Factors</th>
<th>Num repeats</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Harmonic</td>
<td>262</td>
<td>Stretch</td>
<td>0.90, 1.11</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>Harmonic</td>
<td>262</td>
<td>Stretch</td>
<td>0.95, 1.05</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>Harmonic</td>
<td>262</td>
<td>Stretch</td>
<td>0.97, 1.03</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>Harmonic</td>
<td>262</td>
<td>Stretch</td>
<td>0.99, 1.01</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>Pure</td>
<td>262</td>
<td>Stretch</td>
<td>0.95, 1.05</td>
<td>4</td>
</tr>
</tbody>
</table>

Table 11.4 Modified Melodies 2-melody set protocol for normal hearing subjects

The results for harmonic tones are plotted in Figure 11.15. All subjects scored at or near 100% for stretch factors (0.9, 1.11), with the exception of subject NH16. Subject NH16 was additionally tested with the more extreme stretch factors (0.67, 1.5), and scored 15/16 and 16/16 respectively (not shown in Figure 11.15). Thus all subjects exhibited the expected pattern of ceiling scores at extreme stretch factors, decreasing to chance scores as the stretch factors approached 1.0. This pattern is shown more clearly in the group result (Figure 11.16), which was calculated as described in §16.6.2. The circles show the median score (50th percentile), and the error bars indicate the 25th and 75th percentiles. A median score of 100% means that more than half of the subjects scored 100%.

There was a trend for higher scores with compressed intervals (stretch factor < 1) compared to stretched intervals (stretch factor > 1). This is particularly apparent for stretch 0.95 (median score 100%) compared to stretch 1.05 (median score 81%).
Figure 11.15 Individual stretch results, 2-melody set, 262 Hz harmonic tones
The results for the 9 subjects who were tested with both harmonic and pure tones are shown in Figure 11.17. There was a clear trend for lower scores with pure tones. A statistical comparison of the harmonic and pure tone results is given in §11.3.3.
11.3.3 Dependence on harmonic content

Testing with both the 8-melody set and the 2-melody set showed a clear trend for lower scores with pure tones than harmonic tones. A Monte Carlo statistical analysis (described in §16.4.3) of the stretch factors that were tested with both tone types is shown in Table 11.5. For subjects NH01 and NH07, the results were pooled across the 8-melody and 2-melody sets. The mean pure tone scores were lower than the harmonic tone scores for all subjects except NH11 (who obtained equal scores, all near ceiling). The $p$-values are given in the last column of Table 11.5, showing that the difference was statistically significant ($p < 0.05$) in 8 out of 11 subjects.
### Table 11.5: Harmonic vs. pure tones: statistics

<table>
<thead>
<tr>
<th>Subject</th>
<th>Stretch Score</th>
<th>Harmonic Score (%)</th>
<th>Pure Score (%)</th>
<th>Harmonic Diff (%)</th>
<th>Pure Diff (%)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>NH01</td>
<td>0.90</td>
<td>41/44</td>
<td>32/40</td>
<td>93</td>
<td>80</td>
<td>-13</td>
</tr>
<tr>
<td></td>
<td>0.95</td>
<td>64/72</td>
<td>29/48</td>
<td>89</td>
<td>60</td>
<td>-28</td>
</tr>
<tr>
<td></td>
<td>1.05</td>
<td>48/72</td>
<td>23/48</td>
<td>67</td>
<td>48</td>
<td>-19</td>
</tr>
<tr>
<td></td>
<td>1.11</td>
<td>43/44</td>
<td>27/40</td>
<td>98</td>
<td>68</td>
<td>-30</td>
</tr>
<tr>
<td>mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-23</td>
<td>&lt;.001***</td>
</tr>
<tr>
<td>NH02</td>
<td>0.95</td>
<td>16/16</td>
<td>20/24</td>
<td>100</td>
<td>83</td>
<td>-17</td>
</tr>
<tr>
<td></td>
<td>0.97</td>
<td>15/16</td>
<td>10/16</td>
<td>94</td>
<td>63</td>
<td>-31</td>
</tr>
<tr>
<td></td>
<td>1.03</td>
<td>16/16</td>
<td>16/16</td>
<td>100</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1.05</td>
<td>16/16</td>
<td>24/24</td>
<td>100</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-12</td>
<td>&lt;.01**</td>
</tr>
<tr>
<td>NH03</td>
<td>0.90</td>
<td>8/8</td>
<td>16/16</td>
<td>100</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.95</td>
<td>16/16</td>
<td>11/16</td>
<td>100</td>
<td>69</td>
<td>-31</td>
</tr>
<tr>
<td></td>
<td>1.05</td>
<td>14/16</td>
<td>15/16</td>
<td>88</td>
<td>94</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>1.11</td>
<td>8/8</td>
<td>16/16</td>
<td>100</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-6</td>
<td>0.09</td>
</tr>
<tr>
<td>NH04</td>
<td>0.95</td>
<td>16/16</td>
<td>6/8</td>
<td>100</td>
<td>75</td>
<td>-25</td>
</tr>
<tr>
<td></td>
<td>1.05</td>
<td>12/16</td>
<td>1/8</td>
<td>75</td>
<td>13</td>
<td>-63</td>
</tr>
<tr>
<td>mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-44</td>
<td>&lt;.001***</td>
</tr>
<tr>
<td>NH05</td>
<td>0.90</td>
<td>20/24</td>
<td>16/24</td>
<td>83</td>
<td>67</td>
<td>-17</td>
</tr>
<tr>
<td></td>
<td>0.95</td>
<td>9/16</td>
<td>8/16</td>
<td>56</td>
<td>50</td>
<td>-6</td>
</tr>
<tr>
<td></td>
<td>1.05</td>
<td>10/16</td>
<td>8/16</td>
<td>63</td>
<td>50</td>
<td>-13</td>
</tr>
<tr>
<td></td>
<td>1.11</td>
<td>23/24</td>
<td>21/24</td>
<td>96</td>
<td>88</td>
<td>-8</td>
</tr>
<tr>
<td>mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-11</td>
<td>0.06</td>
</tr>
<tr>
<td>NH06</td>
<td>0.95</td>
<td>15/16</td>
<td>9/16</td>
<td>94</td>
<td>56</td>
<td>-38</td>
</tr>
<tr>
<td></td>
<td>1.05</td>
<td>14/16</td>
<td>13/16</td>
<td>88</td>
<td>81</td>
<td>-6</td>
</tr>
<tr>
<td>mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-22</td>
<td>0.02*</td>
</tr>
<tr>
<td>NH07</td>
<td>0.95</td>
<td>32/32</td>
<td>4/16</td>
<td>100</td>
<td>25</td>
<td>-75</td>
</tr>
<tr>
<td></td>
<td>1.05</td>
<td>25/32</td>
<td>13/16</td>
<td>78</td>
<td>81</td>
<td>3</td>
</tr>
<tr>
<td>mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-36</td>
<td>&lt;.001***</td>
</tr>
<tr>
<td>NH08</td>
<td>0.95</td>
<td>16/16</td>
<td>11/16</td>
<td>100</td>
<td>69</td>
<td>-31</td>
</tr>
<tr>
<td></td>
<td>1.05</td>
<td>15/16</td>
<td>14/16</td>
<td>94</td>
<td>88</td>
<td>-6</td>
</tr>
<tr>
<td>mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-19</td>
<td>0.02*</td>
</tr>
<tr>
<td>NH09</td>
<td>0.90</td>
<td>8/8</td>
<td>8/8</td>
<td>100</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.95</td>
<td>8/8</td>
<td>26/32</td>
<td>100</td>
<td>81</td>
<td>-19</td>
</tr>
<tr>
<td></td>
<td>1.05</td>
<td>8/8</td>
<td>21/32</td>
<td>100</td>
<td>66</td>
<td>-34</td>
</tr>
<tr>
<td></td>
<td>1.11</td>
<td>7/8</td>
<td>7/8</td>
<td>88</td>
<td>88</td>
<td>0</td>
</tr>
<tr>
<td>mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-13</td>
<td>0.03*</td>
</tr>
<tr>
<td>NH10</td>
<td>0.95</td>
<td>16/16</td>
<td>13/16</td>
<td>100</td>
<td>81</td>
<td>-19</td>
</tr>
<tr>
<td></td>
<td>1.05</td>
<td>16/16</td>
<td>15/16</td>
<td>100</td>
<td>94</td>
<td>-6</td>
</tr>
<tr>
<td>mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-13</td>
<td>0.03*</td>
</tr>
<tr>
<td>NH11</td>
<td>0.95</td>
<td>7/8</td>
<td>7/8</td>
<td>88</td>
<td>88</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1.05</td>
<td>8/8</td>
<td>8/8</td>
<td>100</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td>0.66</td>
</tr>
</tbody>
</table>
These results are consistent with earlier studies (reviewed in §2.4.4) that showed better fundamental frequency discrimination with harmonic tones compared to pure tones (Henning and Grosberg 1968; Spiegel and Watson 1984; Moore and Peters 1992). Platt and Racine (1985) observed better performance with harmonic tones than pure tones on a pitch-matching task. In a musical context, Trainor (1996) measured the ability of normal hearing subjects to detect a pitch change in a melody (the procedure was described in §5.2.2), and found performance was significantly worse for pure tones (group mean $d' = 0.3$) than for complex tones comprising the first five harmonics with equal amplitudes ($d' = 1.4$). These studies demonstrate that harmonic tones have more salient pitch than pure tones.

### 11.3.4 Dependence on musical experience

Figure 11.18 shows histograms of the scores for harmonic tones with stretch factor 1.05 modification, for 15 adult subjects, classified into two groups according to musical experience (7 musicians and 8 non-musicians, as per Table 11.2). Stretch factor 1.05 was used because it exhibited the largest inter-quartile range in the group results (Figure 11.16), so was potentially a sensitive indicator. The scores were placed into bins by calculating an equivalent score out of 16 trials, rounding where necessary. Perfect scores were obtained by a majority of the musicians, but by none of the non-musicians. The two subject groups have significantly different distributions of scores according to the Wilcoxon rank-sum test ($p = 0.008$).

![Histogram of stretch 1.05 harmonic tone scores for musicians and non-musicians](image)
These results are consistent with previous studies that showed improved performance on pitch-based tasks by musically-trained subjects. Spiegel and Watson (1984) investigated the effect of musical training on frequency discrimination. They measured frequency difference limens ($\Delta f/f$) in the range 0.10 – 0.45% for musicians, and 0.18 – 1.7% for non-musicians. Although the distributions of the two groups overlapped, half of the non-musicians had a threshold exceeding that of the worst musician. Musicians also out-performed non-musicians on a tone pattern discrimination task using ten sequential tones. Platt and Racine (1985) observed better performance by musicians than non-musicians on a pitch-matching task.

### 11.3.5 Nudge results

As will be described in §11.4, the results of testing cochlear implant recipients created a desire for an alternative to the Stretch distortion. The Nudge distortion (§11.2.3) was therefore developed and validated with six normal hearing subjects. As before, harmonic tones with a base frequency of 262 Hz were used, with the default melody set (Old MacDonald and Twinkle Twinkle Little Star). Nudge factors 2, 1, 0.5 and 0.25 were tested. The individual results are plotted in Figure 11.19. Subject NH05, aged 74, had the poorest scores. The remaining subjects (who were aged under 45) obtained perfect scores for a nudge of 0.5 semitones or greater. No subject scored significantly better than chance for a 0.25 semitone nudge.

The poorer result for the oldest subject is consistent with the study by Moore and Peters (1992), reviewed in §2.4.4 (see Figure 2.15), who found that elderly subjects had worse fundamental frequency discrimination than younger subjects, despite near-normal hearing thresholds.

In contrast to the results of Trainor and Trehub (1992) (§11.2.3), there was no evidence that a one semitone change (which went outside the key) was more easily detected than a two-semitone change (which remained within the key), as both these changes resulted in equal scores (at or near ceiling) in all subjects. An important difference in procedure is that Trainor and Trehub (1992) used an unfamiliar melody, whereas the Modified Melodies test used familiar melodies. As discussed in §2.7.2, Dowling and Fujitani (1971) argued that the exact interval sizes of a familiar melody were stored in memory, but only the contour of an unfamiliar melody was initially stored. It would therefore be expected that the central processing for detecting changes would differ between familiar and unfamiliar melodies.
Figure 11.19 Individual nudge results, 262 Hz harmonic tones.

The group median results are shown in Figure 11.20. The implied psychometric function is very steep, with perfect performance for a 0.5 semitone nudge plummeting to chance performance for a 0.25 semitone nudge.
11.3.6 Comparison of scores for each melody

One benefit of having only two melodies in the set is that there are enough trials to compare the scores for each melody. The size of the intervals may influence the ability to detect changes. Twinkle, Twinkle, Little Star contains one large interval (a fifth, 7 semitones), and all other intervals are one or two semitones. Old MacDonald contains two large intervals (5 and 9 semitones). The harmonic tone scores (Figure 11.15) for each melody were summed across stretch factors for each of the 15 subjects, as shown in Table 11.6. The last column of Table 11.6 lists the two-sided $p$-value for the difference between the scores for the two melodies, treating each subject individually (according to the "brute-force" method of §16.2.3). No subject showed a statistically significant difference. The last row of Table 11.6 shows that for the group as a whole, the mean difference between percent-correct scores for the two melodies was only one percentage point, and a paired comparison (according to §16.4.3) indicated no statistically significant difference. A similar analysis of the Nudge results (Figure 11.19) also showed no statistically significant difference between the two melodies (Table 11.7), despite all five subjects having slightly better scores for Old MacDonald.
### Table 11.6 Comparing Stretch scores for the two melodies

<table>
<thead>
<tr>
<th>Subject</th>
<th>Old Mac score</th>
<th>Twinkle score</th>
<th>Old Mac (%)</th>
<th>Twinkle (%)</th>
<th>Diff (%)</th>
<th>2-sided p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>NH01</td>
<td>90/108</td>
<td>82/108</td>
<td>83</td>
<td>76</td>
<td>-7</td>
<td>0.21</td>
</tr>
<tr>
<td>NH02</td>
<td>39/40</td>
<td>40/40</td>
<td>98</td>
<td>100</td>
<td>3</td>
<td>0.54</td>
</tr>
<tr>
<td>NH03</td>
<td>22/24</td>
<td>24/24</td>
<td>92</td>
<td>100</td>
<td>8</td>
<td>0.26</td>
</tr>
<tr>
<td>NH04</td>
<td>21/24</td>
<td>23/24</td>
<td>88</td>
<td>96</td>
<td>8</td>
<td>0.42</td>
</tr>
<tr>
<td>NH05</td>
<td>38/48</td>
<td>39/48</td>
<td>79</td>
<td>81</td>
<td>2</td>
<td>0.90</td>
</tr>
<tr>
<td>NH07</td>
<td>34/36</td>
<td>33/36</td>
<td>94</td>
<td>92</td>
<td>-3</td>
<td>0.81</td>
</tr>
<tr>
<td>NH09</td>
<td>24/24</td>
<td>21/24</td>
<td>100</td>
<td>88</td>
<td>-13</td>
<td>0.13</td>
</tr>
<tr>
<td>NH10</td>
<td>32/32</td>
<td>32/32</td>
<td>100</td>
<td>100</td>
<td>0</td>
<td>1.00</td>
</tr>
<tr>
<td>NH11</td>
<td>11/12</td>
<td>12/12</td>
<td>92</td>
<td>100</td>
<td>8</td>
<td>0.54</td>
</tr>
<tr>
<td>NH12</td>
<td>41/48</td>
<td>40/48</td>
<td>85</td>
<td>83</td>
<td>-2</td>
<td>0.89</td>
</tr>
<tr>
<td>NH13</td>
<td>58/64</td>
<td>51/64</td>
<td>91</td>
<td>80</td>
<td>-11</td>
<td>0.11</td>
</tr>
<tr>
<td>NH14</td>
<td>49/64</td>
<td>53/64</td>
<td>77</td>
<td>83</td>
<td>6</td>
<td>0.44</td>
</tr>
<tr>
<td>NH15</td>
<td>48/64</td>
<td>53/64</td>
<td>75</td>
<td>83</td>
<td>8</td>
<td>0.33</td>
</tr>
<tr>
<td>NH16</td>
<td>45/64</td>
<td>51/64</td>
<td>70</td>
<td>80</td>
<td>9</td>
<td>0.26</td>
</tr>
<tr>
<td>NH18</td>
<td>14/16</td>
<td>14/16</td>
<td>88</td>
<td>88</td>
<td>0</td>
<td>1.00</td>
</tr>
<tr>
<td>mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>0.52</td>
</tr>
</tbody>
</table>

### Table 11.7 Comparing Nudge scores for the two melodies

<table>
<thead>
<tr>
<th>Subject</th>
<th>Old Mac score</th>
<th>Twinkle score</th>
<th>Old Mac (%)</th>
<th>Twinkle (%)</th>
<th>Diff (%)</th>
<th>2-sided p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>NH01</td>
<td>54/62</td>
<td>51/62</td>
<td>87</td>
<td>82</td>
<td>-5</td>
<td>0.53</td>
</tr>
<tr>
<td>NH02</td>
<td>7/12</td>
<td>5/12</td>
<td>58</td>
<td>42</td>
<td>-17</td>
<td>0.54</td>
</tr>
<tr>
<td>NH05</td>
<td>31/39</td>
<td>30/39</td>
<td>79</td>
<td>77</td>
<td>-3</td>
<td>0.89</td>
</tr>
<tr>
<td>NH08</td>
<td>19/23</td>
<td>18/23</td>
<td>83</td>
<td>78</td>
<td>-4</td>
<td>0.85</td>
</tr>
<tr>
<td>NH17</td>
<td>21/30</td>
<td>20/30</td>
<td>70</td>
<td>67</td>
<td>-3</td>
<td>0.89</td>
</tr>
<tr>
<td>mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-6</td>
<td>0.27</td>
</tr>
</tbody>
</table>

### 11.4 Cochlear implant recipients

Following the pilot testing with normal hearing subjects (§11.3.1), testing of cochlear implant recipients commenced with subject S06 (Table 15.1), using the 8-melody set shown in Figure 11.13. Subject S06 was one of the best performers in the pitch ranking experiments of chapters 9 and 10, so his results would provide guidance as to whether the test was appropriate for cochlear implant recipients generally. After scoring at chance for stretch factors (0.5, 2.0), the subject admitted that his memory of some of the melodies was uncertain. Informal melody identification tests and discussions with other recipients suggested that, despite the names of the melodies being very well known to the recipients, the melodies themselves were not familiar enough for the purposes of the test. It therefore seemed prudent to revise the test procedure, and subsequent testing used the two-melody set of Old MacDonald and Twinkle Twinkle Little Star.
11.4.1 Results

Figure 11.21 shows the results for 13 cochlear implant recipients, three in Sydney (S01, S06 and S08 from Table 15.1) and ten in Melbourne (M01 – M10). The Melbourne results were collected by Stewart (2006) as part of a study on the benefit of contralateral hearing aids (§11.5), but the results in this section are for cochlear implant alone. The melodies were presented as harmonic tones with a base frequency of 262 Hz (C4, as used with the normal-hearing subjects in §11.3.2) from a loudspeaker in a sound-treated room. Subjects listened with their usual processor. The protocol is more fully described in §11.5, but briefly, the subjects were first tested with reciprocal stretch factors distant from 1.0, and then the stretch factors were moved closer to 1.0 until chance scores were obtained.

It is immediately apparent that the cochlear implant recipients performed much worse than the normal-hearing group (compare to Figure 11.15 and Figure 11.16; note that the stretch factors used for implant recipients covered a much larger range than those used for normal hearing subjects). For stretch factors 0.90 and 1.11, most normal hearing subjects obtained perfect scores, but the cochlear implant subjects scored at chance levels (most cochlear implant subjects were not tested at those stretch factors because their performance had already fallen to chance at more extreme stretch factors). Only six cochlear implant subjects (M05, M06, M07, M10, S01, S06) obtained scores significantly better than chance for stretch factor 2.0; yet to a normal hearing listener, a stretch of 2.0 is so extreme that the melody is almost unrecognisable. The remaining seven cochlear implant subjects apparently obtained very little pitch information at all. It is likely that subjects M02, M03, M04, M09, and S08 could only reliably detect that the stretch factor 4 melodies were different to all the others. Subject M01 scored significantly above chance only at stretch factor 0.25. Subject M08 did not score better than chance at any stretch factor.

In many subjects, the expected pattern of increasing scores at more extreme intervals was not seen for compressed intervals (stretch factor < 1). This is apparent in the group median scores (Figure 11.22; note that to avoid bias, the missing data points for the poorer performers were estimated by interpolation, as described in §16.6.2). The median score at stretch factor 0.25 is no better than at stretch factor 0.50, and is in stark contrast to the 100% median score at stretch factor 4.0. The inter-quartile ranges indicated in Figure 11.22 demonstrate a much greater variability in scores across cochlear implant recipients than was observed with normal hearing subjects (Figure 11.16). The largest inter-quartile spread in scores occurred for stretch factor 2.0.
Figure 11.21 CI individual stretch results, 262 Hz harmonic tones
11.5 Benefit of contralateral hearing aid

The Modified Melodies test was used in a University of Melbourne Master of Audiology thesis by Penny Stewart (Stewart 2006), supervised by Cathy Sucher, to investigate the benefit to cochlear implant recipients of a contralateral hearing aid. Details of the ten Melbourne subjects are listed in Table 15.1. Each subject regularly wore a contralateral hearing aid. Performance was tested under two conditions: cochlear implant alone (CI), and cochlear implant in conjunction with a contralateral hearing aid (Bimodal). In addition, Sydney subject S08, who also regularly wore a contralateral hearing aid, was tested with the same protocol.

The test protocol was devised in collaboration with the present author. The default melodies were used (Old MacDonald and Twinkle Twinkle Little Star), presented as harmonic tones with base frequency 262 Hz (C4). Subjects were initially tested with four blocks having the reciprocal pair of stretch factors (0.5, 2.0), with the order balanced across the two conditions as shown in Table 11.8, giving 16 trials for each stretch factor. If those scores were not significantly above chance (or if time permitted), the next four blocks were performed with stretch factors (0.25, 4.0) in the same balanced order of conditions. Subsequently, for each condition (CI or Bimodal), the stretch factors were adjusted to be closer to 1.0 until chance scores were obtained.
<table>
<thead>
<tr>
<th>Block</th>
<th>Condition</th>
<th>Tone</th>
<th>Base Freq (Hz)</th>
<th>Distortion</th>
<th>Factors</th>
<th>Number of trials</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CI</td>
<td>Harmonic</td>
<td>262</td>
<td>Stretch</td>
<td>0.5, 2.0</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>Bimodal</td>
<td>Harmonic</td>
<td>262</td>
<td>Stretch</td>
<td>0.5, 2.0</td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td>Bimodal</td>
<td>Harmonic</td>
<td>262</td>
<td>Stretch</td>
<td>0.5, 2.0</td>
<td>8</td>
</tr>
<tr>
<td>4</td>
<td>CI</td>
<td>Harmonic</td>
<td>262</td>
<td>Stretch</td>
<td>0.25, 4.0</td>
<td>8</td>
</tr>
<tr>
<td>5</td>
<td>Bimodal</td>
<td>Harmonic</td>
<td>262</td>
<td>Stretch</td>
<td>0.25, 4.0</td>
<td>8</td>
</tr>
<tr>
<td>6</td>
<td>Bimodal</td>
<td>Harmonic</td>
<td>262</td>
<td>Stretch</td>
<td>0.25, 4.0</td>
<td>8</td>
</tr>
<tr>
<td>7</td>
<td>Bimodal</td>
<td>Harmonic</td>
<td>262</td>
<td>Stretch</td>
<td>0.25, 4.0</td>
<td>8</td>
</tr>
<tr>
<td>8</td>
<td>CI</td>
<td>Harmonic</td>
<td>262</td>
<td>Stretch</td>
<td>0.25, 4.0</td>
<td>8</td>
</tr>
</tbody>
</table>

Table 11.8 Modified Melodies protocol for CI v bimodal study

11.5.1 Results

The results for the bimodal condition are listed in Table 11.9 and plotted in Figure 11.23, with the cochlear implant alone results (from Figure 11.21) repeated for comparison. The large improvements for the bimodal condition are evident in the group median results (Figure 11.24). The bimodal condition had the expected pattern of increasing scores for more extreme stretches, with a median score of 100% for stretch factor 0.25, unlike the cochlear implant alone condition.

Although the median bimodal score was much better than the median cochlear implant score, it was still markedly worse than the median normal hearing score. For stretch factors (0.90, 1.11), the median bimodal score was close to chance (Figure 11.24), in contrast to the median normal hearing score of 100% (Figure 11.16). It should be remembered that these subjects had severe to profound hearing loss in their contralateral ear, which has been found to degrade pitch perception (Moore and Carlyon 2005).

To aid comparison of the two conditions for individual subjects, summary scores were calculated for each subject (according to §16.6.1), for stretches between 0.50 and 2.0 (not all subjects and conditions were tested at stretch factors 0.25 and 4.0). A scatter plot of the bimodal summary score against the cochlear implant summary score (Figure 11.25) revealed that all subjects except M07 obtained better scores in the bimodal condition. A statistical analysis (according to §16.4.3) of the raw scores for those stretch factors that were common to both conditions (including stretch factors 0.25 and 4.0 where possible, as listed in Table 11.9) showed that the improvement was statistically significant in 7 out of 11 subjects (M01, M02, M03, M05, M08, M09, and S08). The significantly-improved subjects included all but one of seven poorest-performing subjects in the cochlear implant alone condition; the remaining poor performer was M04, whose bimodal improvement was almost significant ($p = 0.08$). Subjects M01, M03, M08, and M09, who are clustered together on the scatter plot (Figure 11.25), mostly scored at chance with cochlear implant alone, but were among the best performers in the bimodal condition. This implies that their poor performance in the cochlear implant alone condition was not merely due to misunderstanding the task, or a poor recollection of the
The Modified Melodies test

<table>
<thead>
<tr>
<th>Subject</th>
<th>Stretch</th>
<th>CI score</th>
<th>Bimodal score</th>
<th>CI %</th>
<th>Bimodal %</th>
<th>Bimodal - CI %</th>
<th>Significance (p value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M01</td>
<td>0.50</td>
<td>10/16</td>
<td>16/16</td>
<td>63</td>
<td>100</td>
<td>38</td>
<td>&lt;.001***</td>
</tr>
<tr>
<td></td>
<td>0.67</td>
<td>9/16</td>
<td>16/16</td>
<td>56</td>
<td>100</td>
<td>44</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1.49</td>
<td>6/16</td>
<td>15/16</td>
<td>38</td>
<td>94</td>
<td>56</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.00</td>
<td>6/16</td>
<td>16/16</td>
<td>38</td>
<td>100</td>
<td>63</td>
<td></td>
</tr>
<tr>
<td></td>
<td>mean</td>
<td></td>
<td></td>
<td>50</td>
<td></td>
<td></td>
<td>&lt;.001***</td>
</tr>
<tr>
<td>M02</td>
<td>0.25</td>
<td>6/16</td>
<td>13/16</td>
<td>38</td>
<td>81</td>
<td>44</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>9/16</td>
<td>7/16</td>
<td>56</td>
<td>44</td>
<td>-13</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.00</td>
<td>9/16</td>
<td>14/16</td>
<td>56</td>
<td>88</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.00</td>
<td>16/16</td>
<td>16/16</td>
<td>100</td>
<td>100</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>mean</td>
<td></td>
<td></td>
<td>16</td>
<td></td>
<td>0.02 *</td>
<td></td>
</tr>
<tr>
<td>M03</td>
<td>0.50</td>
<td>12/16</td>
<td>15/16</td>
<td>75</td>
<td>94</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.67</td>
<td>6/16</td>
<td>14/16</td>
<td>38</td>
<td>88</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.49</td>
<td>6/16</td>
<td>16/16</td>
<td>38</td>
<td>100</td>
<td>63</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.00</td>
<td>9/16</td>
<td>16/16</td>
<td>56</td>
<td>100</td>
<td>44</td>
<td></td>
</tr>
<tr>
<td></td>
<td>mean</td>
<td></td>
<td></td>
<td>44</td>
<td></td>
<td>&lt;.001***</td>
<td></td>
</tr>
<tr>
<td>M04</td>
<td>0.25</td>
<td>12/16</td>
<td>16/16</td>
<td>75</td>
<td>100</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>13/16</td>
<td>15/16</td>
<td>81</td>
<td>94</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.67</td>
<td>11/16</td>
<td>11/16</td>
<td>69</td>
<td>69</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.49</td>
<td>5/16</td>
<td>7/16</td>
<td>31</td>
<td>44</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.00</td>
<td>9/16</td>
<td>9/16</td>
<td>56</td>
<td>56</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.00</td>
<td>16/16</td>
<td>16/16</td>
<td>100</td>
<td>100</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>mean</td>
<td></td>
<td></td>
<td>8</td>
<td></td>
<td>0.08</td>
<td></td>
</tr>
<tr>
<td>M05</td>
<td>0.50</td>
<td>5/8</td>
<td>16/16</td>
<td>63</td>
<td>100</td>
<td>38</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.67</td>
<td>5/16</td>
<td>16/16</td>
<td>31</td>
<td>100</td>
<td>69</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.90</td>
<td>11/16</td>
<td>5/8</td>
<td>69</td>
<td>63</td>
<td>-6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.11</td>
<td>10/16</td>
<td>7/8</td>
<td>63</td>
<td>88</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.49</td>
<td>13/16</td>
<td>16/16</td>
<td>81</td>
<td>100</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.00</td>
<td>8/8</td>
<td>16/16</td>
<td>100</td>
<td>100</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>mean</td>
<td></td>
<td></td>
<td>24</td>
<td></td>
<td>&lt;.001***</td>
<td></td>
</tr>
<tr>
<td>M06</td>
<td>0.50</td>
<td>15/16</td>
<td>13/16</td>
<td>94</td>
<td>81</td>
<td>-13</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.67</td>
<td>8/16</td>
<td>7/16</td>
<td>50</td>
<td>44</td>
<td>-6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.90</td>
<td>8/16</td>
<td>7/16</td>
<td>50</td>
<td>44</td>
<td>-6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.11</td>
<td>6/16</td>
<td>9/16</td>
<td>38</td>
<td>56</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.49</td>
<td>7/16</td>
<td>13/16</td>
<td>44</td>
<td>81</td>
<td>38</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.00</td>
<td>14/16</td>
<td>13/16</td>
<td>88</td>
<td>81</td>
<td>-6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>mean</td>
<td></td>
<td></td>
<td>4</td>
<td></td>
<td>0.29</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>0.50</td>
<td>0.67</td>
<td>0.90</td>
<td>1.11</td>
<td>1.49</td>
<td>2.00</td>
</tr>
<tr>
<td>-------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td></td>
<td>8/16</td>
<td>7/16</td>
<td>10/16</td>
<td>7/16</td>
<td>9/16</td>
<td>11/16</td>
<td>16/16</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>44</td>
<td>63</td>
<td>44</td>
<td>56</td>
<td>69</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>69</td>
<td>13</td>
<td>50</td>
<td>50</td>
<td>56</td>
<td>69</td>
<td>81</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>-31</td>
<td>-13</td>
<td>6</td>
<td>0</td>
<td>-6</td>
<td>-19</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M08</td>
<td>0.50</td>
<td>11/16</td>
<td>16/16</td>
<td>69</td>
<td>100</td>
<td>69</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.67</td>
<td>8/16</td>
<td>11/16</td>
<td>50</td>
<td>69</td>
<td>50</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.49</td>
<td>6/16</td>
<td>15/16</td>
<td>38</td>
<td>94</td>
<td>38</td>
<td>94</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.00</td>
<td>2/16</td>
<td>16/16</td>
<td>13</td>
<td>100</td>
<td>13</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S08</td>
<td>0.25</td>
<td>10/16</td>
<td>15/16</td>
<td>63</td>
<td>94</td>
<td>63</td>
<td>94</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>12/16</td>
<td>16/16</td>
<td>75</td>
<td>100</td>
<td>75</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.67</td>
<td>13/16</td>
<td>10/16</td>
<td>81</td>
<td>63</td>
<td>81</td>
<td>63</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.50</td>
<td>10/16</td>
<td>16/16</td>
<td>63</td>
<td>100</td>
<td>63</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.00</td>
<td>10/16</td>
<td>16/16</td>
<td>63</td>
<td>100</td>
<td>63</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.00</td>
<td>16/16</td>
<td>16/16</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 11.9 Modified Melodies Cl v Bimodal results
Figure 11.23 Individual stretch results: CI alone vs. bimodal, 262 Hz harmonic tones
Figure 11.24 Group median results: CI alone and Bimodal, 262 Hz harmonic tones

Figure 11.25 Scatter plot of summary scores of Bimodal vs. CI alone
11.5.2 Relationship to contralateral hearing loss

It was hypothesised that the benefit obtained from a contralateral hearing aid would depend on the degree of hearing loss in the contralateral ear, i.e. better contralateral hearing should provide more benefit. The benefit was quantified as the summary score of the bimodal condition, minus the summary score for the cochlear implant alone. Figure 11.26 shows the unaided audiograms of the contralateral ears of the Melbourne subjects. As expected for cochlear implant recipients, most had moderate to severe hearing loss at frequencies below 1000 Hz, and profound loss at higher frequencies. For the purpose of analysis, the thresholds were averaged over the frequency range up to 1000 Hz. This matched the spectral profile of the harmonic tones, which fell off above 1000 Hz (§11.2.1, Figure 11.4). Scatter plots of the two statistically significant relationships are shown in Figure 11.27. In support of the hypothesis, the benefit of the contralateral hearing aid was significantly negatively correlated with hearing thresholds ($r = -0.75, p = 0.01$; top panel of Figure 11.27). The subject with the worst contralateral hearing, M07, was the only subject who had a negative bimodal benefit. The cluster of subjects (M01, M03, M08 and M09) identified in the earlier scatter plot (Figure 11.25) had the most contralateral residual hearing (lowest thresholds), and obtained the most benefit (also see Table 11.9).

The correlation between the bimodal scores and contralateral hearing thresholds was not significant ($r = 0.46, p = 0.18$). This was somewhat unexpected, since the bimodal benefit was correlated. Instead, the scores for cochlear implant alone were very highly correlated with the contralateral hearing thresholds ($r = 0.93, p = 1e-4$; bottom panel of Figure 11.27). In other words, the subjects with the most contralateral hearing (M01, M03, M08 and M09) obtained the worst scores with their implant alone, and this largely explains why they obtained the most bimodal benefit. Perhaps these subjects had learnt to rely on the relatively good pitch perception provided by their contralateral hearing aid, and were unaccustomed to the pitch sensation provided by the implant alone.

These subjects had all used hearing aids in the period before cochlear implantation, and had decided to be implanted because they were no longer obtaining sufficient benefit from their hearing aids. Yet they had all persisted with a contralateral hearing aid after implantation. The research question was: how much benefit did the contralateral hearing aid provide? Because the hearing aid alone condition was not tested, this study could not indicate whether there was any synergy between the cochlear implant and the hearing aid. However, as is clear from the scatter plot (Figure 11.25), there was no correlation between the summary scores for the cochlear implant alone and bimodal conditions. Given the poor scores observed for cochlear implant alone, it seems likely that the subjects were primarily relying on their hearing aids in the bimodal condition. This is consistent with the closed-set familiar melody identification study by
Kong et al. (2005), which was reviewed in §6.6. In that study, the four regular users of contralateral hearing aids had scores in the bimodal condition that were significantly higher than with cochlear implant alone, but were not significantly different from their scores with hearing aid alone. Kong et al. concluded that in the bimodal condition, the subjects were not combining pitch information from the two ears, but were instead primarily attending to the better (hearing aid) ear.
Figure 11.26 Contralateral ear audiograms
11.6 Limitations of stretch distortion

The results in §11.4 and §11.5 showed that the Modified Melodies test was suitable for use with cochlear implant recipients, and that it was sensitive enough to demonstrate the benefit of a contralateral hearing aid.

However, the study exposed some limitations of the Stretch distortion when used with large stretch factors. Figure 11.28 is a plot of the notes of the two original melodies and the corresponding modified versions with stretch factor 4. The first issue is that stretching affected the two melodies differently. The first note of each melody was kept unchanged, and the remaining notes were placed using stretched intervals from that reference point. Because Twinkle Twinkle Little Star begins with its lowest note (C), the stretched melody extended above the original frequency range. Conversely, because Old MacDonald begins with F, the stretched melody extended both above and below the original frequency range. Thus the stretch factor 4 version of Twinkle Twinkle Little Star occupied a frequency range more than an octave higher than the corresponding modified version of Old MacDonald. In contrast, for normal hearing subjects, the largest stretch factor routinely used was 1.11, and the two melodies...
differed in range by less than a semitone (compare Figure 11.28 to Figure 11.5 and Figure 11.6).

Of more concern is that the increased frequency range may have introduced unintended cues. The harmonic tone stimuli had been designed with the intention of removing timbre cues by keeping the spectral profile constant as fundamental frequency varied (Figure 11.4). However, for a base frequency of C4 (262 Hz), the highest note of Twinkle Twinkle Little Star with stretch factor 2 was F#5 (740 Hz). For stretch factor 4, the highest note was C7 (2093 Hz), which was beyond the band-pass filter upper corner frequency of 1000 Hz. These high notes were still audible because the tones were all normalised to have equal amplitude, but their timbre was markedly different to that of the original melodies.

Electroodograms for the original and modified melodies, with stretch factors 0.25, 0.50, 2.0 and
The melodies were processed by ACE (8 maxima, 22 channels, 1800 pps). The amplitude modulations in the envelopes are not visible at this time scale. The electroodograms suggest that the cochlear implant recipients may have utilised a gross place cue: the modified melodies with stretch factors 2 and 4 excited a different set of electrodes to the original melodies. Conversely, the modified melodies with stretch factors less than one excited the same set of electrodes as the originals. This is the most likely reason for the shape of the cochlear implant group median results (Figure 11.22), where stretch factor 0.25 had a much lower score than stretch factor 4. It also explains why subjects M02, M03, M04, M09, and S08 could only reliably detect that the stretch factor 4 melodies were different to all the others.

It could be argued that with stretch factor 2 or 4, the test was no longer a melody perception test, but effectively an electrode discrimination task. This undermines the primary motivation for developing the Modified Melodies test, which was to test whether the subject was actually perceiving melody, rather than brightness changes (§11.1). Therefore, it is recommended that stretch factors in future studies not exceed 1.33 (which expands the 9-semitone range of these melodies to a more moderate one octave). Stretch factors less than 1.33 are well suited for normal hearing subjects, but will often be too difficult for cochlear implant recipients. It was this observation that prompted the development of the Nudge distortion type, where the modified melody can be constrained to span the same frequency range as the original melody.
Figure 11.29 Spectrograms of Old MacDonald
Figure 11.30 Spectrograms of Twinkle Twinkle Little Star
11.7 Conclusion

This chapter has described the development of the new Modified Melodies test. Although it was designed for testing cochlear implant recipients, it appears to be a sensitive test for exploring aspects of pitch perception in normal hearing. Significantly higher scores were obtained with harmonic tones than pure tones. Subjects with at least five years musical experience performed significantly better than those without.

Cochlear implant recipients performed much worse than subjects with normal hearing, and many of them could not extract any useful pitch information from harmonic tones in the octave beginning at middle C (a very common frequency range). A contralateral hearing aid provided a large benefit. The subjects with the most contralateral hearing had the worst scores with their implant alone. Bimodal performance was still markedly worse than normal hearing performance.
12  HWR strategy take-home study

12.1  Introduction

The ACE strategy uses quadrature envelope detection at the output of each filter (§4.4.3). An alternative is to use a half-wave rectifier (HWR), with little or no smoothing (§6.2.1). The HWR acts as a simple model of the Inner Hair Cells in the normal cochlea (§2.4.2), retaining the temporal fine structure. As reviewed in Chapter 6, this approach has been used clinically by several implant manufacturers: the Phase-Locked CIS strategy for the Laura cochlear implant (Peeters et al. 1993), the HiRes strategy from Advanced Bionics (Wilson et al. 2004; Wilson et al. 2005), and the FSP strategy for the MED-EL Pulsar implant (Hochmair et al. 2006; Arnoldner et al. 2007). Despite this, surprisingly little research has been carried out on the effects of half-wave rectification on pitch perception.

Although HWR showed little benefit in the sung-vowel pitch-ranking study in Chapter 10, the cochlear implant recipients had very limited exposure to the research strategies, as the MATLAB implementation precluded take-home use. It was possible that it would take time for a recipient to learn to use the enhanced temporal cues. This chapter describes the implementation and take-home trial of the HWR strategy on the Freedom processor.

12.2  DSP implementation

12.2.1  Freedom integrated circuit

The Freedom processor (Figure 3.1) is based upon a single custom integrated circuit (Figure 12.1) containing analog circuitry, a microcontroller and four identical DSP cores (Swanson et al. 2007). There are three 16-bit sigma-delta Analog-to-Digital Converters (ADCs), and a 16-bit Digital-to-Analog Converter (DAC). The microcontroller manages the user interface, responding to button presses and updating the LCD. It transfers code and data from an external Flash memory into the DSP memories at power-up, or when the recipient requests a change of program. It monitors the battery voltage and alerts the recipient when it is running low. It communicates with a PC over a serial port when the processor is being fitted to a recipient.

Each DSP core has a 1024-word program memory, and three 1024-word data memories denoted X, Y and Z. Each core has multiple execution units operating in parallel. Each instruction is 128 bits wide with separate fields for each execution unit. In one cycle, a core can execute two multiply-accumulates, two arithmetic-logic unit (ALU) operations, and five memory load or store operations with address register updates (the X and Y memories can be accessed twice in each processor cycle). The data path is 16 bits wide. The DSPs use a 16-bit High Speed Bus to
communicate with each other, and with the peripherals.

![Image of Freedom processor integrated circuit](image)

**Figure 12.1 Freedom processor integrated circuit**

### 12.2.2 Freedom processor ACE implementation

Figure 12.2 is a block diagram of the ACE strategy on the Freedom processor, showing how the processing is divided across the four DSP cores (Swanson et al. 2007).

The Freedom processor has both a directional and an omni-directional microphone. On DSP1, the two microphone signals can be combined by a beam-forming algorithm to produce a directional response. The algorithm adapts to the sound environment and steers a null in the response pattern towards the loudest noise that is not in front of the recipient (Spriet et al. 2007). An auxiliary audio signal (from a built-in telecoil, or an assistive listening device such as an FM receiver, or a consumer audio device) can also be mixed into the signal path.

The Gain Control block on DSP2 implements manual sensitivity control and fast acting AGC. It optionally incorporates the Whisper feature (syllabic compression, McDermott et al. 2002), or a slow-acting Auto-sensitivity function (Patrick et al. 2006). DSP2 also provides an audio output for monitor headphones, typically used by the parents of a child recipient.

DSP3 implements the 128-point FFT vector-sum filterbank (§4.5). After the filterbank, the gain of each channel can be adjusted to shape the overall frequency response according to recipient preference. Alternatively, Adaptive Dynamic Range Optimization (ADRO) can be applied (James et al. 2002). This is a form of slow acting gain control, which operates on each band independently. It estimates the long-term amplitude statistics of each filter envelope, and adjusts the gain so that each frequency band is presented at a comfortable level.

DSP4 implements maxima selection (§4.7.2) and amplitude mapping (§4.8).
12.2.3 Freedom processor HWR implementation

To implement HWR, the code on DSP3 was modified at the point following the vector sum of the FFT bins (Equation 4.21). For the low-frequency channels, the real component was half-wave-rectified. For the remaining channels, the usual quadrature envelope detection (QED) was calculated, as shown in Figure 12.3. The number of channels that were half-wave-rectified was

![Figure 12.2 Freedom processor software architecture](image-url)
HWR strategy take-home study

controlled by a map parameter. As the clinical fitting software had no knowledge of this new parameter, the author wrote a python script that could modify it in the processor memory map. A value of zero gave an ACE map, and a non-zero value gave a HWR map.

A value of zero gave an ACE map, and a non-zero value gave a HWR map.

Each filter output will contain components that have frequencies lying in the pass-band of that filter. The unsmoothed HWR signal thus provides temporal cues at frequencies that lie in the pass-band of the filter. As the first four filters in the default ACE frequency allocation are centred at 250, 375, 500, and 625 Hz, only the two lowest frequency channels were half-wave rectified. There was little reason to apply half-wave rectification to the higher frequency channels, as subjects were unlikely to perceive temporal cues at frequencies of 500 Hz or higher. Note that the same reasoning was applied in the design of the Phase-Locked CIS strategy (Peeters et al. 1993) (§6.2.1), and the FSP strategy (Hochmair et al. 2006; Arnoldner et al. 2007) (§6.3.6), where only the two or three lowest frequency channels were half-wave rectified. The stimulation rate of the HWR strategy needs to be high enough to adequately sample the waveforms of the half-wave rectified filter outputs. The clinical fitting software limits the choice (in the interests of simplicity) to a handful of rates: 250, 500, 720, 900, 1200, 1800, 2400 and 3500 pps. In line with the earlier streaming experiments (chapters 9 and 10), a stimulation rate of 1800 pps was chosen, allowing 8 maxima to be selected for Nucleus 24 subjects, which gives a total stimulation rate of 14400 pps. Considering the typical temporal pitch saturation limit of 300 Hz (§3.7), the next lower clinical rate (1200 pps) only barely satisfies the criterion of being at least four times the modulation frequency (McKay et al. 1994). The next higher clinical rate of 2400 pps only allows 6 maxima for Nucleus 24 subjects, which was considered inadequate.

Figure 12.4 contrasts the electrodograms of ACE and HWR for a 262 Hz (C4) harmonic tone (the spectrogram of this tone was shown in Figure 4.12). As described in §4.6, channels 1, 3, and 5 each contain only a single harmonic, and so ACE (using quadrature envelope detection) has no modulation on those channels. The other channels each contain two harmonics, and so the quadrature envelope modulates at F0. In the HWR strategy, channel 1 is driven by the half-wave-rectified first harmonic and so modulates at F0. Channel 2 is driven by the half-wave-rectified first harmonic and so modulates at F0.
rectified mixture of the first and second harmonics, so the modulation is less regular. Another feature of the HWR strategy is that the maxima selection process is given the opportunity to select the lower-amplitude channels 9 and 10 during the "off-time" of channels 1 and 2. Compared to ACE, the HWR strategy appears to enhance the temporal cues.

![Figure 12.4](image_url)

Figure 12.4 Electrograms of ACE and HWR for a 262 Hz harmonic tone. Only the 12 lowest frequency channels (out of 22) are shown.

### 12.3 Methods

In comparing half-wave rectification to envelope detection, it was important not to change other
aspects of the strategy, in particular, stimulation rate. It would not be a fair comparison to trial HWR at 1800 pps against ACE at 900 pps, as the increased stimulation rate in itself could affect performance. A higher rate could potentially represent amplitude modulation cues more faithfully (McKay et al. 1994). Conversely, there is evidence that sensitivity to temporal modulation is worse at higher rates (Galvin and Fu 2005). The ideal subject for a comparison was therefore someone who regularly used ACE at 1800 pps. Unfortunately, only two such subjects were available in Sydney, and none in Melbourne. For subjects who used lower rates, the protocol was designed so that they were initially fit with ACE at 1800 pps (denoted ACE1800), then fit with HWR at 1800 pps (HWR1800), and then the two new maps were compared in a repeated-measures design.

12.3.1 Sydney pilot study

The purpose of the pilot study was to gain confidence that the new HWR strategy would provide acceptable sound quality before asking a larger group of subjects to wear it exclusively for periods of several weeks (§12.3.2). The participants were two Nucleus 24 recipients, S01 and S06, and one Freedom implant recipient, S09 (Table 15.1). Subjects S01 and S06 had shown good sensitivity to amplitude modulation cues in previous experiments (Chapters 9 and 10).

Subject S06 was a long-time user of ACE at 1800 pps, so a balanced protocol was not undertaken. Instead, he was fitted with HWR at 1800 pps in the first session, and evaluated with both strategies in that initial session and at three additional sessions at one week intervals. Subject S09 also regularly used ACE at 1800 pps. He was fitted with HWR at 1800 pps in the first session, and evaluated with both strategies in that initial session and in one additional session two weeks later. These two subjects had both ACE and HWR strategies in their Freedom processor for the entire study duration, and were instructed to switch between them in different environments and compare sound quality.

Subject S01 usually used 900 pps ACE, and had negligible experience with 1800 pps stimulation outside of lab-based testing (chapters 8 and 10). Therefore the protocol shown in Figure 12.5 was followed. ACE at 1800 pps was fitted in the first session, and used for one week. In the next session, HWR was fitted, and the subject was instructed to use both strategies for the following three weeks, switching between them to compare sound quality. In the final two phases, each strategy was worn alone for three weeks. Speech and pitch perception were measured with both strategies at weeks 1, 3, 6, and 9.
12.3.2 Melbourne study

The protocol for the Melbourne study was designed by Dr Pam Dawson, and the study was carried out by Dr Dawson and Michelle Knight at Cochlear Ltd, Melbourne. The protocol is summarised in Figure 12.6.

Ten adult Nucleus 24 recipients were recruited for the study. All had at least two years experience with their implant. Before entering the study, all subjects were using ACE with stimulation rates of 900 pps or lower, in line with standard practice at the Melbourne cochlear implant clinic. The subjects were fitted with 1800 pps ACE in the first session, and given four weeks of take-home experience to become accustomed to it. Two subjects had difficulty adjusting to the higher rate and chose to withdraw from the study at that point.

The remainder of the study was an ABAB comparison of the ACE and HWR strategies. The strategy that was started at week four was counterbalanced across the group, to avoid bias due to learning effects (Figure 12.6). A strategy was used for a four week interval, after which speech and pitch perception with that strategy were evaluated. In the final two-week phase, both ACE and HWR programs were loaded into the subject's processor. The programs were identified to the subject only by their processor slot number (i.e. "P1" and "P2").

Two further subjects withdrew from the study for personal reasons before completing the full protocol. Details of the six subjects who completed the protocol are given in Table 15.1.
<table>
<thead>
<tr>
<th>Week</th>
<th>Use ACE1800</th>
<th>Fit ACE1800</th>
<th>Evaluate HWR1800</th>
<th>Use HWR1800</th>
<th>Evaluate ACE1800</th>
<th>Use both</th>
<th>Questionnaire</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Use ACE1800</td>
<td>Fit ACE1800</td>
<td>Use HWR1800</td>
<td>Use HWR1800</td>
<td>Evaluate ACE1800</td>
<td>Use both</td>
<td>Questionnaire</td>
</tr>
<tr>
<td>4</td>
<td>Use ACE1800</td>
<td>Fit HWR1800</td>
<td>Evaluate HWR1800</td>
<td>Use HWR1800</td>
<td>Evaluate ACE1800</td>
<td>Use both</td>
<td>Questionnaire</td>
</tr>
<tr>
<td>8</td>
<td>Use HWR1800</td>
<td>Evaluate HWR1800</td>
<td>Use HWR1800</td>
<td>Use HWR1800</td>
<td>Evaluate ACE1800</td>
<td>Use both</td>
<td>Questionnaire</td>
</tr>
<tr>
<td>12</td>
<td>Use HWR1800</td>
<td>Evaluate ACE1800</td>
<td>Use HWR1800</td>
<td>Use HWR1800</td>
<td>Evaluate ACE1800</td>
<td>Use both</td>
<td>Questionnaire</td>
</tr>
<tr>
<td>16</td>
<td>Use HWR1800</td>
<td>Evaluate HWR1800</td>
<td>Use HWR1800</td>
<td>Use HWR1800</td>
<td>Evaluate ACE1800</td>
<td>Use both</td>
<td>Questionnaire</td>
</tr>
<tr>
<td>20</td>
<td>Use HWR1800</td>
<td>Evaluate ACE1800</td>
<td>Use HWR1800</td>
<td>Use HWR1800</td>
<td>Evaluate ACE1800</td>
<td>Use both</td>
<td>Questionnaire</td>
</tr>
<tr>
<td>22</td>
<td>Use HWR1800</td>
<td>Evaluate HWR1800</td>
<td>Use HWR1800</td>
<td>Use HWR1800</td>
<td>Evaluate ACE1800</td>
<td>Use both</td>
<td>Questionnaire</td>
</tr>
</tbody>
</table>

Figure 12.6 Melbourne ACE v HWR study protocol.

At week 4, subjects were randomly assigned to one of the two protocols shown.

12.3.3 Fitting

Because the two most apical channels were half wave rectified, it was expected that this could reduce their loudness. The overall loudness was checked informally by talking with the recipient. If the recipient felt that the HWR program was softer, then both the threshold and comfort levels on the two apical channels were increased by 5% of their dynamic range. The remaining channels had the same current levels as in the recipient's ACE program.

12.3.4 Speech perception

Because the HWR strategy differed from ACE only in the two lowest frequency channels, and these channels make relatively small contributions to speech intelligibility, no improvement was expected for speech perception in quiet. Still, it was important to verify that there was no degradation; as mentioned in §6.4.1, Green, Faulkner et al. (2005) found worse vowel perception for their F0 Modulation strategy compared to CIS.

Speech perception was evaluated with CNC words in quiet at 60 dB SPL. Each list contained 50 words. For Sydney subjects S01 and S06, both strategies were tested in each session, using either one or two lists each (depending on time), with the testing order balanced across sessions. Speech perception was not tested with S09. In the Melbourne protocol, one strategy was tested per session, using three word lists.

12.3.5 Pitch perception

Pitch perception was evaluated with the Modified Melodies test, using harmonic tones. In chapter 11, the original melodies spanned the frequency range of C4 – A4 (262 – 440 Hz).
Although typical of music, this was not an optimal frequency range for cochlear implant temporal pitch perception. As the HWR strategy was intended to enhance temporal cues, testing at a lower frequency range was more appropriate. A base frequency of 149 Hz was chosen so that the melodies were kept within the frequency range where temporal cues were most salient (i.e. the original melodies spanned the range 149 – 250 Hz).

Sydney subjects used the default melodies (Old MacDonald and Twinkle Twinkle Little Star). Subjects S01 and S06 used the Stretch distortion, with stretch factors limited to the range (0.75, 1.33) as recommended in §11.6. Subject S09 was tested with the Nudge distortion using a nudge of 2, 4, and 6 semitones.

The Modified Melodies protocol used in the Melbourne study is summarised in Table 12.1. Subjects began with the Ascending steps test. This was a variant of the Modified Melodies test where the correct melody was an ascending sequence of seven notes, two semitones apart, covering a one octave range (149 – 297 Hz), as shown in black in Figure 12.7. The distorted version had the middle note shifted down by either two, four or six semitones. The subjects also performed the Modified Melodies test using the default melodies, with nudge factors 2, 4, and 6. Prior to each block, a short practice run was performed using a nudge of 6 semitones, with the researcher providing verbal feedback.

The Ascending melody was expected to yield higher scores than the default melodies for three reasons: firstly, it was much simpler than a typical melody, having an identical interval repeated six times; secondly, the same melody appeared in every trial (instead of alternating between two different melodies); and thirdly, the modification disturbed the contour, whereas the nudge modifications in Old MacDonald and Twinkle Twinkle Little Star changed the size of the intervals but preserved the contour.

<table>
<thead>
<tr>
<th>Melody</th>
<th>Tone</th>
<th>Distortion</th>
<th>Factor</th>
<th>Number of repeats</th>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ascending</td>
<td>Harmonic</td>
<td>Nudge</td>
<td>6</td>
<td>4</td>
<td>Practice</td>
</tr>
<tr>
<td>Ascending</td>
<td>Harmonic</td>
<td>Nudge</td>
<td>2, 4, 6</td>
<td>8</td>
<td>Test</td>
</tr>
<tr>
<td>Old Mac, Twinkle</td>
<td>Harmonic</td>
<td>Nudge</td>
<td>6</td>
<td>4</td>
<td>Practice</td>
</tr>
<tr>
<td>Old Mac, Twinkle</td>
<td>Harmonic</td>
<td>Nudge</td>
<td>2, 4, 6</td>
<td>8</td>
<td>Test</td>
</tr>
</tbody>
</table>

Table 12.1 Modified Melodies protocol for Melbourne ACE v HWR study
Figure 12.7. Ascending melody (black), and Nudge 2 (red), Nudge 4 (blue) and Nudge 6 (pink) distortions

12.4 Results

12.4.1 Speech perception

Figure 12.8 shows the mean scores, pooled across test sessions, for CNC word scores in quiet for the ACE and HWR strategies for each subject. The error bars show the standard deviation across lists. The number of lists performed for each strategy was eight for subject S01, five for subject S06, and six for subjects M11 – M15. The right-most bars show the group mean with error bars indicating one standard deviation.

Statistical analysis according to the resampling method (Simon 1997) revealed no significant difference between the ACE and HWR strategies for any individual subject, nor for the group result ($p = 0.24$, two-sided test).

A more traditional statistical analysis was performed by Dr Dawson on the Melbourne results. A two-way repeated measures analysis of variance (ANOVA) was used to determine the effect of the factors “strategy” and “session” on the arcsin of the average score for the three lists in each session. There were no significant differences for the group between the ACE and HWR strategies ($F(1,5) = 1.28$, $p = 0.31$). The main effect of “session” was not significant ($F(1,5) = 1.63$, $p = 0.25$). Independent t-tests for each individual subject found no significant differences
between the two strategies for any subject.

![Figure 12.8 CNC word scores ACE v HWR](image)

### 12.4.2 Melody perception

The results and statistics (analysed according to §16.4.3) for subjects S01 and S06, pooled across sessions, are listed in Table 12.2, and plotted in Figure 12.9. Neither subject showed a significant difference between the two strategies. Subject S01 showed a practice effect with both strategies, being tested only at stretches (0.75, 1.33) in week 1, and scoring better with stretches (0.80, 1.25) in later weeks.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Stretch</th>
<th>ACE score</th>
<th>HWR score</th>
<th>ACE (%)</th>
<th>HWR (%)</th>
<th>HWR – ACE (%)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>S01</td>
<td>0.75</td>
<td>13/16</td>
<td>13/16</td>
<td>81</td>
<td>81</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.80</td>
<td>13/24</td>
<td>14/24</td>
<td>88</td>
<td>58</td>
<td>-29</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td>0.90</td>
<td>9/16</td>
<td>8/16</td>
<td>56</td>
<td>50</td>
<td>-6</td>
<td>0.6</td>
</tr>
<tr>
<td></td>
<td>1.11</td>
<td>14/16</td>
<td>13/16</td>
<td>88</td>
<td>81</td>
<td>-6</td>
<td>0.6</td>
</tr>
<tr>
<td></td>
<td>1.25</td>
<td>24/24</td>
<td>24/24</td>
<td>100</td>
<td>100</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1.33</td>
<td>13/16</td>
<td>13/16</td>
<td>81</td>
<td>81</td>
<td>0</td>
<td>0.91</td>
</tr>
<tr>
<td>mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-7</td>
<td>0.23</td>
</tr>
</tbody>
</table>

| S06      | 0.75    | 16/24     | 14/24     | 67      | 58      | -8            | 0.8     |
|          | 0.80    | 13/24     | 18/24     | 54      | 75      | 21            | 0.23    |
|          | 1.25    | 17/24     | 19/24     | 71      | 79      | 8             | 0.05    |
|          | 1.33    | 21/24     | 21/24     | 88      | 88      | 0             | 0.00    |
| mean     |         |           |           |         |         | 5             | 0.23    |

Table 12.2 Modified Melodies Stretch results, 149 Hz harmonic tones
Figure 12.9 Modified Melodies Stretch results, 149 Hz harmonic tones
The results of the Melbourne subjects for the Ascending melody are listed in Table 12.3, and plotted in Figure 12.10. Contrary to expectations, subjects found this task very difficult. Almost all individual scores were at chance levels (50%), and the expected pattern of increasing scores for larger nudges was not seen in individual results or in the group median result (Figure 12.11). There was no significant difference between ACE and HWR.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Nudge</th>
<th>ACE score</th>
<th>HWR score</th>
<th>ACE (%)</th>
<th>HWR (%)</th>
<th>HWR - ACE (%)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>M11</td>
<td>2.00</td>
<td>12/16</td>
<td>15/16</td>
<td>75</td>
<td>94</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.00</td>
<td>6/16</td>
<td>10/16</td>
<td>38</td>
<td>63</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6.00</td>
<td>11/16</td>
<td>6/16</td>
<td>69</td>
<td>38</td>
<td>-31</td>
<td>0.37</td>
</tr>
<tr>
<td>mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M12</td>
<td>2.00</td>
<td>9/16</td>
<td>10/16</td>
<td>56</td>
<td>63</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.00</td>
<td>5/16</td>
<td>7/16</td>
<td>31</td>
<td>44</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6.00</td>
<td>8/16</td>
<td>4/16</td>
<td>50</td>
<td>25</td>
<td>-25</td>
<td>0.62</td>
</tr>
<tr>
<td>mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-2</td>
<td></td>
</tr>
<tr>
<td>M13</td>
<td>2.00</td>
<td>8/16</td>
<td>3/16</td>
<td>50</td>
<td>19</td>
<td>-31</td>
<td>0.63</td>
</tr>
<tr>
<td></td>
<td>4.00</td>
<td>5/16</td>
<td>5/16</td>
<td>31</td>
<td>31</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6.00</td>
<td>9/16</td>
<td>13/16</td>
<td>56</td>
<td>81</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-10</td>
<td>0.91</td>
</tr>
<tr>
<td>M14</td>
<td>2.00</td>
<td>5/16</td>
<td>6/16</td>
<td>31</td>
<td>38</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.00</td>
<td>2/16</td>
<td>3/16</td>
<td>13</td>
<td>19</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6.00</td>
<td>7/16</td>
<td>0/16</td>
<td>44</td>
<td>0</td>
<td>-44</td>
<td>0.10</td>
</tr>
<tr>
<td>mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-17</td>
<td>0.93</td>
</tr>
<tr>
<td>M15</td>
<td>2.00</td>
<td>2/8</td>
<td>8/16</td>
<td>25</td>
<td>50</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.00</td>
<td>3/8</td>
<td>7/16</td>
<td>38</td>
<td>44</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6.00</td>
<td>4/8</td>
<td>11/16</td>
<td>50</td>
<td>69</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>17</td>
<td>0.10</td>
</tr>
<tr>
<td>M16</td>
<td>2.00</td>
<td>6/8</td>
<td>9/16</td>
<td>75</td>
<td>56</td>
<td>-19</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.00</td>
<td>5/8</td>
<td>5/16</td>
<td>63</td>
<td>31</td>
<td>-31</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6.00</td>
<td>6/8</td>
<td>12/16</td>
<td>75</td>
<td>75</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-17</td>
<td>0.93</td>
</tr>
</tbody>
</table>

Table 12.3 Ascending Nudge results, 149 Hz harmonic tones
Figure 12.10 Ascending Nudge results, 149 Hz harmonic tones
The results of the Melbourne subjects and Sydney subject S09 with the default melodies are listed in Table 12.4, and plotted in Figure 12.12. Two subjects (M12 and M14) scored much better with the default melodies than with the Ascending melody. Subject M12 showed the expected pattern of increasing scores for larger nudges with HWR. Subject M14 scored near ceiling with the default melodies for both strategies. He had obtained scores consistently below 50% with the Ascending melody, including a score of 0/16 for nudge 6 with HWR; it is possible that he misunderstood the Ascending task and was answering opposite to that requested. The remaining subjects scored at levels near chance. No subject showed a significant difference between ACE and HWR. The group median results (Figure 12.13) show a slight trend for better performance with ACE.
<table>
<thead>
<tr>
<th>Subject</th>
<th>Nudge</th>
<th>ACE score</th>
<th>HWR score</th>
<th>ACE (%)</th>
<th>HWR (%)</th>
<th>HWR - ACE (%)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>S09</td>
<td>2.00</td>
<td>7/16</td>
<td>11/16</td>
<td>44</td>
<td>69</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.00</td>
<td>18/24</td>
<td>11/16</td>
<td>75</td>
<td>69</td>
<td>-6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6.00</td>
<td>9/16</td>
<td>9/16</td>
<td>56</td>
<td>56</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6</td>
<td>0.27</td>
</tr>
<tr>
<td>M11</td>
<td>2.00</td>
<td>19/32</td>
<td>15/32</td>
<td>59</td>
<td>47</td>
<td>-13</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.00</td>
<td>19/32</td>
<td>18/32</td>
<td>59</td>
<td>56</td>
<td>-3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6.00</td>
<td>22/32</td>
<td>21/32</td>
<td>69</td>
<td>66</td>
<td>-3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-6</td>
<td>0.83</td>
</tr>
<tr>
<td>M12</td>
<td>2.00</td>
<td>14/24</td>
<td>18/32</td>
<td>58</td>
<td>56</td>
<td>-2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.00</td>
<td>20/24</td>
<td>26/32</td>
<td>83</td>
<td>81</td>
<td>-2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6.00</td>
<td>18/24</td>
<td>32/32</td>
<td>75</td>
<td>100</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td></td>
<td>mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7</td>
<td>0.14</td>
</tr>
<tr>
<td>M13</td>
<td>2.00</td>
<td>18/32</td>
<td>17/32</td>
<td>56</td>
<td>53</td>
<td>-3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.00</td>
<td>21/32</td>
<td>17/32</td>
<td>66</td>
<td>53</td>
<td>-13</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6.00</td>
<td>18/32</td>
<td>19/32</td>
<td>56</td>
<td>59</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-4</td>
<td>0.74</td>
</tr>
<tr>
<td>M14</td>
<td>2.00</td>
<td>16/16</td>
<td>27/32</td>
<td>100</td>
<td>84</td>
<td>-16</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.00</td>
<td>16/16</td>
<td>31/32</td>
<td>100</td>
<td>97</td>
<td>-3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6.00</td>
<td>14/16</td>
<td>32/32</td>
<td>88</td>
<td>100</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td></td>
<td>mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-2</td>
<td>0.74</td>
</tr>
<tr>
<td>M15</td>
<td>2.00</td>
<td>8/16</td>
<td>16/32</td>
<td>50</td>
<td>50</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.00</td>
<td>6/16</td>
<td>17/32</td>
<td>38</td>
<td>53</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6.00</td>
<td>11/16</td>
<td>14/32</td>
<td>69</td>
<td>44</td>
<td>-25</td>
<td></td>
</tr>
<tr>
<td></td>
<td>mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-3</td>
<td>0.66</td>
</tr>
<tr>
<td>M16</td>
<td>2.00</td>
<td>10/16</td>
<td>19/32</td>
<td>63</td>
<td>59</td>
<td>-3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.00</td>
<td>9/16</td>
<td>13/32</td>
<td>56</td>
<td>41</td>
<td>-16</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6.00</td>
<td>8/16</td>
<td>15/32</td>
<td>50</td>
<td>47</td>
<td>-3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-7</td>
<td>0.81</td>
</tr>
</tbody>
</table>

Table 12.4 Modified Melodies Nudge results, 149 Hz harmonic tones
Figure 12.12 Modified Melodies Nudge results, 149 Hz harmonic tones
12.5 Conclusion

A total of nine subjects had take-home experience with the HWR strategy on the Freedom processor. No significant difference in speech perception or melody perception was found between the HWR and ACE strategies. A possible explanation is that the better-performing subjects (e.g. S01, S06, M12, M14) were capable of extracting the temporal modulation cues that were present on many channels in the ACE strategy (Figure 12.4). The enhanced modulation that the HWR strategy produced on the two most apical channels did not make these temporal cues more salient. The poorer-performing subjects were apparently unable to extract temporal cues from either strategy.

These results can be compared to those of Arnoldner et al. (2007) for the FSP strategy, which also used half-wave rectification on the two most apical channels (§6.3.6). They found worse performance on a melody discrimination test with FSP than with CIS.

Regarding test procedures, subjects found the Ascending test more difficult than the default melodies. This was possibly because it was not a familiar melody, so subjects had no internal template to compare against, or perhaps the intervals (two semitones) were too small. Further use of the Ascending test is not recommended. The Exchange and Backwards distortion types were subsequently developed to provide an easier task than the Nudge distortion.
13 Melody perception with place cues

13.1 Introduction

In their recent review of cochlear implant pitch perception, Moore and Carlyon (2005) argued that "a demonstration that pitch can be conveyed purely by place-of-excitation cues would disprove models which propose that timing cues are essential for pitch perception". The experiment in this chapter tested the hypothesis that cochlear implant place cues alone are sufficient to convey a melody.

13.2 Methods and stimuli

One apparent difficulty in conveying a melody with only place pitch cues is the relatively large spacing of the electrodes. For example, the Nucleus straight electrode array has an electrode spacing of 0.75 mm, which according to Greenwood (1990) represents a musical interval of about two semitones. In the frequency allocation typically used by Nucleus implant recipients with the SPrint or Freedom processors, the three most-apical electrodes are assigned to the centre frequencies 250, 375 and 500 Hz, spanning an entire octave. As mentioned in §5.3, Chen and Zeng (2005) tackled this difficulty by assigning groups of successive notes (either 3 or 6 successive semitones) to the same electrode, but this is hardly an ideal solution.

A much better approach is to use the intermediate place-pitch percepts that are generated by stimulating more than one electrode concurrently. This could be done (as in §8.3) by synthesising dual-electrode pulse trains (McDermott and McKay 1994; Kwon and van den Honert 2006). An even simpler method is to apply pure tones to the ACE strategy, as was demonstrated in §8.4. Indeed, this is so straightforward that it is surprising that it has not been attempted earlier.

The electrodogram in Figure 13.1 shows the Old MacDonald melody played using pure tones with the ACE strategy, with the lowest note being C5 (523 Hz), and the highest A5 (880 Hz). The stimulation pattern for each note has no amplitude modulation, so that only place cues are available. Each note activates multiple electrodes, due to the broad analysis filters in the processor. Notes C5 and D5 excite the same electrodes, but with differing amplitudes. In such cases, the centroid of the stimulation pattern may provide a place-pitch cue, as described in §8.5.
Previous research, reviewed in §3.8.1, supported the view that cochlear implant place-pitch cues were more akin to the brightness attribute of timbre than to pitch. It was therefore expected that cochlear implant recipients would find melody perception with place cues alone to be difficult or even impossible. For this reason, the Modified Melodies test protocol for this study was designed to be as easy as possible. Some cochlear implant recipients in the previous studies (§11.4, §11.5 and §12) had reported anecdotally that having to switch their attention from trial to trial between Old MacDonald and Twinkle Twinkle Little Star was distracting, and that it would be easier to concentrate if the same melody was used in every trial. Therefore a single melody (Old MacDonald) was used in the protocol.

As before, in each trial the original and a modified version of the melody were presented in random order, and the subject had to select the original version. Each block of trials contained four types of pitch modification, in randomised order, as shown in Figure 13.2. The original melody is shown in the bottom panel. To briefly recap the descriptions from §11.2, in the Nudge modifications, the 5th and 6th notes (DD) were shifted up by either two semitones ("N2") or five semitones ("N5"). In the Exchange ("Ex") modification, the 5th and 6th notes (DD) were exchanged with the 10th and 11th notes (GG). In the Backwards ("Bk") modification, the melody was reversed in time. The Exchange and Backwards modifications disturb the melodic contour, while the Nudge modifications preserve the contour but change the size of the musical intervals.

Each modified melody had the same frequency range (i.e. the same highest and lowest notes) as the original, to prevent subjects using the type of unintentional cues which probably occurred with large stretches (as explained in §11.6). Another potential source of non-pitch cues was inherent in the design of the test, in that the original version of the melody appeared in every trial, paired with a modified version which varied from trial to trial. Thus it was possible that a
subject could learn to recognise the specific, unchanging pattern of stimulation produced by the original melody, without necessarily perceiving its pitch. For example, imagine that note D5 had some characteristic non-pitch quality that the subject could easily identify; perhaps a sensation of roughness, or even merely of being louder than the neighbouring notes. Then a subject might identify the original melody purely by the presence of this distinctive note: as can be seen in Figure 13.2, note D5 is not contained at all in the N2 and N5 modifications, and detecting its appearance at the wrong time in the Ex and Bk modifications does not (in this hypothetical example) require a sense of pitch. Such behaviour had been considered unlikely when there were multiple melodies in the set, but it was more of a risk when the same melody was used on each trial. To avoid this, a new version of the software was created that incorporated random transposition. The original melody was transposed up by 0, 1, 2 or 3 semitones (selected randomly on each trial), so that it produced a slightly different stimulation pattern on each trial. The modified melody in each trial was also transposed up by the same amount as the original.

![Figure 13.2 Old MacDonald, original and modified melodies](image)

Seven cochlear implant recipients participated in the experiment: S01, S02, S06, S08, S09, S10, and S11 (Table 15.1). All subjects used their usual map on their own Freedom processor, with a variety of stimulation rates: 250 pps for S02; 900 pps for S01, S08, S10, S11; and 1800 pps for
S06 and S09. The SPEAK strategy (used by subject S02) has an average stimulation rate of 250 pps, but the period between stimulation pulses is jittered to avoid any rate-pitch percept. The pure tones had sufficiently high frequency (≥ 523 Hz) that the envelopes had no amplitude modulation, as verified by inspection of electrodograms; thus there were no temporal cues to the pitch of the stimuli.

The test protocol is summarised in Table 13.1. Each block contained 16 trials (1 melody × 4 modifications × 4 repeats). Subjects performed three or four blocks (12 or 16 trials per modification).

<table>
<thead>
<tr>
<th>Tone</th>
<th>Base Freq (Hz)</th>
<th>Modifications</th>
<th>Num repeats</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pure</td>
<td>523</td>
<td>N2 N5 Ex Bk</td>
<td>4</td>
</tr>
</tbody>
</table>

Table 13.1 Modified Melodies protocol for place cues study

### 13.3 Results

Percent-correct scores for the four modifications, for each subject, are shown in Figure 13.3 (with error bars indicating 90% confidence intervals). Scores significantly above chance were obtained by all subjects for the Backwards modification; by all subjects except S02 for the Exchange modification, by four subjects (S01, S06, S09, and S10) for the N5 modification; and by only one subject (S10) for the N2 modification. Subject S02, the only subject who used 250 pps stimulation rate, had the poorest results, only scoring above chance for the backwards modification. The group median results are shown in Figure 13.4 (with error bars indicating the interquartile range). The intended gradation in difficulty from Backwards to Nudge 2 was achieved.

The possibility that subjects were recognising patterns of brightness changes cannot be completely ruled out. However, anecdotal reports from subjects suggested that they were hearing a melody based on pitch changes. Subjects S01, S06, S08, and S09 had participated in earlier experiments using the Modified Melodies test (see Table 15.2). When asked to describe how well they could hear the melody in this experiment, compared to previous experiments, they each said simply that the instrument had changed. For example, subject S06 described the 149 Hz harmonic tones (§12.4.2) as a trombone, and these 523 Hz pure tones as a pan flute. A proposal for future research is to investigate the performance of normal hearing subjects on this test when the "melodies" are conveyed by brightness changes.
Figure 13.3 Individual results: place cues only (523 Hz pure tones)
13.4 Conclusion

Using place cues alone, all the subjects in this study were able to recognise an incorrect melodic contour (Backwards modification). Four out of seven subjects were able to detect a large error in the size of a musical interval (N5 modification), and one subject was able to detect a smaller error (N2 modification). This is evidence that not only can the sensation of cochlear implant place-pitch be ordered along a low-to-high scale, but that musical intervals and melodies can be perceived. Place-pitch therefore meets the operational definition of pitch given in §2.2: "that attribute of sensation whose variation is associated with musical melodies" (Plack and Oxenham 2005a).

The inability of the majority of subjects to recognise the two-semitone change in interval may simply reflect the size of their difference limens for place-pitch. Another issue is that the frequency allocated by the sound processor to an electrode does not match the characteristic frequency corresponding to that electrode position. This frequency-to-position mismatch is likely to distort the representation of musical intervals. Subjects sometimes commented that neither presentation of the melody sounded correct.

The finding that place cues alone are sufficient to convey a melody presents a challenge to models of pitch perception in normal hearing that rely on temporal cues (Moore and Carlyon 2005). This issue will be taken up in the next chapter.
14 Conclusions

It's life Jim, but not as we know it.
– Dr McCoy, Star Trek

It's sound Jim, but not as we know it.
– Wai-Kong Lai, Melbourne 1987

When talking to a cochlear implant recipient, it is often possible to forget that they are deaf. This is a remarkable tribute to the success of the technology. However, their hearing deficit quickly becomes apparent in a noisy situation, or when they confess their disappointment with music. Sitting with recipients through countless distorted presentations of Old MacDonald builds a profound awareness of the dismal quality of cochlear implant pitch perception.

This final chapter firstly summarises the experimental results of this thesis chapter by chapter, then attempts to put them into context among the existing body of knowledge on cochlear implant pitch perception. It proposes a model of pitch perception that can account for both normal hearing and cochlear implant results, at least in a qualitative way. It concludes by considering the implications for cochlear implant sound processing.

14.1 Summary of experimental results

Chapter 8 explored place pitch perception. Sequential stimulation of neighbouring channels produced intermediate place pitch percepts. The currents required to equalize the loudness of the dual-channel stimuli were consistent with the McKay loudness model (McKay et al. 2003). Intermediate place pitch percepts were also produced by pure tones processed by the ACE strategy. The centroid model of place pitch perception (Laneau et al. 2004) was extended by incorporating the McKay loudness model to provide a better fit to the experimental results.

Chapter 9 explored temporal pitch perception with single-channel stimuli. A high-rate pulse train, modulated on and off at frequency F0, had higher pitch than a train of pulses at the rate of F0. If amplitude modulation of high-rate pulse trains was to be used to convey pitch, then the shape of the modulating waveform was important: the half-wave shape was better than the square-wave (on-off) shape. There were large differences between recipients in their sensitivity to pulse timing variations: the just noticeable timing deviation ranged from 120 µs to 740 µs. The experimental results were all consistent with a model that derived pitch from the times between successive auditory nerve firings (first-order inter-spike intervals). The results were not consistent with the auto-correlation (all-order inter-spike intervals) model.
Chapter 10 described experiments where both place and temporal cues to pitch were available. For pure tone stimuli, both place cues and temporal cues were useful indicators of fundamental frequency. However, there were few instances of subjects additively combining the two cues; some subjects appeared to ignore the place cue in favour of the temporal cue. For the sung vowel stimuli, place cues were sometimes misleading, and pitch reversals could be predicted by the place-pitch centroid model. The best-performing subjects were those who ignored place cues. Those subjects who performed poorly with ACE obtained some improvement from the HWR, TPS, and F0M research strategies, which enhanced temporal cues.

Chapter 11 described the development of the new Modified Melodies test. Normal hearing listeners had higher scores with harmonic tones than pure tones. Musicians performed better than non-musicians. Cochlear implant recipients performed much worse than subjects with normal hearing. A contralateral hearing aid provided a large benefit.

Chapter 12 described the implementation and take-home trial of the HWR strategy on the Nucleus Freedom processor. There was no significant difference in speech perception or melody perception between the HWR and ACE strategies.

Chapter 13 contained the important finding that cochlear implant place cues did support melody perception.

### 14.2 Cochlear implant place and temporal cues

To understand the principles of cochlear implant perception, researchers synthesize sequences of pulses where a single physical parameter of the stimulus varies in a controlled manner.

Temporal cues can be investigated by stimulating a single electrode with a varying pulse rate. Cochlear implant recipients can recognise melodies presented in this way for pulse rates less than 300 Hz (Pijl and Schwarz 1995; Pijl 1997a). Similar percepts are produced by varying the modulation frequency of an amplitude-modulated high rate pulse train. Thus temporal cues satisfy the definition of pitch.

Place cues can be investigated by choosing a fixed pulse rate, high enough to avoid temporal cues (e.g. 1800 pps), and then varying the electrode position. When a group of neighbouring electrodes are stimulated, the place percept depends on the centroid of the stimulation pattern (§3.6.2)(Laneau et al. 2004), as demonstrated by the pitch-ranking results for pure tones (§8.5) and sung vowels (§10.3.3). This thesis provides the first good evidence that cochlear implant recipients can recognise melodies presented by varying the centroid of the stimulation pattern (§13), so place cues also satisfy the definition of pitch.

The ability of place cues to support melodies may not surprise many workers in the cochlear
implant field; after all, the percept is conventionally called "place pitch". However, the evidence remains that pulse rate and electrode position form two independent perceptual dimensions (§3.8.1) (Tong et al. 1983a; McKay et al. 2000). Place cues depend on which neurons are firing, while temporal cues depend on firing times. This raises a puzzling question: if place and rate are different percepts, based on such dissimilar neural attributes, how can they both be pitch?

The resolution of this dilemma was alluded to in the opening quote of this chapter. Putting it bluntly, cochlear implant recipients do not have normal hearing. The percepts generated by electrical stimulation may have no exact counterparts in normal hearing.

Hearing evolved to provide an organism with information about its surroundings. Pitch perception allows two simultaneous sounds, occupying overlapping spectral regions, to be segregated into two perceptually distinct sound sources, based on their fundamental frequencies. Regarding pitch strength, resolved frequency components evoke a strong pitch (as long as the frequency is below 4000 Hz), and unresolved components evoke a weak pitch. This can be explained in evolutionary terms: sounds with resolved harmonics are most common in the natural environment. Sounds consisting of only unresolved harmonics, or sounds with very high fundamental, are rarely heard, and so the ability to extract pitch from them has little impact on survival. The central pitch processor is adapted to respond most strongly to the auditory nerve firing patterns produced by the most common tonal sounds.

A cochlear implant can produce auditory nerve firing patterns that are unlike any that occur in a normal hearing ear. With such an aberrant input, the central pitch processor cannot be expected to produce a normal response. Such stimulation may induce a partial or distorted response in the central pitch processor, sufficient to allow performance on pitch and melody tests that is above chance, but at a lower level than that achieved with normal hearing. In summary, it is not surprising that cochlear implant recipients perform worse than normal hearing subjects on pitch perception tasks. The following sections address the issue of what specific mechanisms are involved, and what can be done to improve performance.

### 14.3 A new pitch perception model

A good pitch perception model must account for cochlear implant pitch perception as well as normal hearing. Researchers are often forced to resort to esoteric acoustic stimuli to probe the finer nuances of pitch perception in normal hearing. Cochlear implants offer an alternative tool, allowing models to be tested in ways that are not possible with normal hearing subjects. For example, with acoustic stimuli, low frequency temporal cues can be produced at basal places of the cochlea (by band-pass filtering a click train, or amplitude modulating a high-frequency carrier), but it is impossible to produce high frequency temporal cues at apical places. In a cochlear implant, place and temporal cues can be manipulated completely independently.
Proponents of the auto-correlation model argue that the pitch of both resolved and unresolved harmonics can be explained by a single mechanism (§2.4.5) (Meddis and O'Mard 1997). However, the auto-correlation model is not consistent with the results of this thesis for single-channel MPP sequences (§9.5), nor with previously published results for single-channel stimuli with alternating time intervals (§3.7.1) (Carlyon et al. 2002; van Wieringen et al. 2003; Carlyon et al. 2008). These experimental results are consistent with a model that analyses first-order inter-spike intervals (Moore 1997), as shown in Figure 14.1a. Despite this success with single channel stimuli, any model that analyses firing patterns at each place independently, and then combines the information across place, cannot explain why the pitch of a low-rate electrical pulse train does not have the quality of an acoustic pure tone. Nerve firing is better synchronised to a 200 pps electrical pulse train than to a 200 Hz acoustic pure tone; analysis of inter-spike intervals predicts that a cochlear implant recipient should perceive a strong pitch. Modifying the temporal model to incorporate place information, by constraining the range of inter-spike intervals examined at each characteristic frequency (Moore 1997; Bernstein and Oxenham 2005) does not improve matters: it predicts that a low-rate electrical pulse train should have a strong pitch only if the rate matches the electrode position. On the contrary, the pitch is largely unaffected by electrode position (§3.8.1) (Pijl 1997b).

An alternative hypothesis is that a specific phase relationship between the nerve firing times at nearby places is required to evoke a strong pitch (§2.4.6). A new model that incorporates this hypothesis is shown in Figure 14.1b. Like the Moore model, this is only a schematic or qualitative model; developing a quantitative computational model is outside the scope of this thesis. The new feature is a neural processing stage, sensitive to local phase relationships, that takes place before the first-order inter-spike interval analysis. The lateral inhibitory networks (LIN) proposed by Shamma (1985b) have the required type of behaviour (§2.4.6). Auditory nerve firing that exhibits a rapid phase transition across a local region of the cochlea (due to a resolved frequency component) produces a strong response. With appropriate weighting parameters, neural firing that is aligned in phase across a local region (due to a group of unresolved harmonics, or due to low-rate electrical stimulation of a single electrode) is partially (but not completely) suppressed. In both cases, the firing times at the input of the LIN are preserved at the output, allowing subsequent inter-spike interval analysis.
14.4 Place pitch and melody

The ability to convey a melody by varying the centroid of the electrical stimulation pattern is an enigma not explained by the proposed pitch model (Figure 14.1b), or indeed any temporal model. In the experiment of chapter 13, there were no temporal cues to pitch whatsoever. Obtaining a pitch sensation in the absence of temporal cues has little counterpart in acoustic hearing. Pure tones above 4000 Hz provide place cues without temporal cues, but do not support melody perception (§2.4.3) (Attneave and Olson 1971). A sound component that excites a distinct place in the apical to mid region of the cochlea is always accompanied by corresponding temporal cues: if it has a bandwidth narrow enough to only excite a localised region of the cochlea, then it must resemble a sinusoid, and so will provide temporal cues.

A possible instance of pitch without temporal cues in normal hearing is found with low-pass filtered noise. Small and Daniloff (1967) presented subjects with a reference noise band, and asked them to adjust the cut-off frequency of a second noise band so that its pitch was an octave higher. The resulting cut-off frequency was close to twice the reference cut-off frequency, and subjects' accuracy was similar to their accuracy in adjusting the frequency of a pure tone to be an octave higher than a reference pure tone. They argued that the ability to judge an octave relationship implied a true pitch sensation, rather than simply an adjustment of brightness (a more convincing demonstration would be to use the Modified Melodies test). If this sensation is
pitch, it is difficult to explain with a temporal model, or with a place model that utilises peaks in the excitation pattern.

In normal hearing, place-of-excitation has multiple roles aside from pitch. It seems likely that a central representation is formed of neural activity as a function of place, i.e. a spectral profile or "internal spectrum". The perception of loudness involves the summation of this internal spectrum across place (Moore 1997, chapter 2). Brightness depends on the centroid of the internal spectrum (§2.5). Vowel formants correspond to the locations of peaks in the internal spectrum (§2.6). Level differences in the internal spectrum between left and right ears contribute to localisation. A cochlear implant sound processing strategy such as CIS or ACE appears to produce an internal spectrum that has comparable characteristics: it allows good vowel perception in quiet, and bilateral cochlear implant recipients can make use of inter-aural level differences for localisation.

Fundamental frequency is perceptually independent of the internal spectrum; for example the identity of the /i/ vowel in the sung vowel stimuli used in §10.3 is unaffected by changes in F0. Yet despite this independence, there must be a close association between the higher level processing of internal spectrum and fundamental frequency. Many cues can be used in the segregation of sound into multiple streams, such as common onsets, common offsets, common modulations, and spatial location (Moore 1997, chapter 7). The key point is that when attention is directed to one particular stream, that stream has its own F0 and its own internal spectrum. In speech perception, a voice pitch difference can be used to segregate and identify two simultaneous voices (§2.6). In attending to one speaker's voice, the pitch is relatively constant, and changes in the internal spectrum (e.g. formant patterns) of that stream convey the phonemes. In music perception, fundamental frequency and internal spectrum are also associated together, but play different roles. For example, when listening to a small ensemble of musicians, it is possible to pick out one instrument and perceive its pitch, loudness and timbre separately from the rest. In the stream for one instrument, timbre is relatively constant, and the pitch variations can carry a melody line.

Thus fundamental frequency and internal spectrum are coupled together in a perceptual stream in a manner that allows variations in either one to be tracked, depending on the context. It therefore seems possible that the part of the brain that recognises melodies has access to both the fundamental frequency and the internal spectrum. Music typically consists of harmonic tones, and in this case variations in fundamental frequency completely dominate the perception of melody. However, in the absence of any periodicity in the stimulus, (such as unmodulated noise bands in acoustic hearing, or high-rate electrical pulse trains in a cochlear implant), the perceptual stream will not have a fundamental frequency associated with it. In this atypical case, it seems that the melody recognition process can make use of variations in the internal spectrum.
14.5 Implications for cochlear implant sound processing

In normal hearing, pitch is dominated by the low-numbered harmonics, which are resolved in the auditory filterbank (§2.4.2). In contrast, the filterbanks typically used in cochlear implant processors do not fully resolve any harmonics for fundamental frequencies in the voice pitch range (§4.6). It is a simple engineering task to build a higher resolution filterbank; for example to use a 512-point FFT instead of the 128-point FFT used in the SPrint or Freedom processor (§4.5). However, if each filter is assigned to one electrode, then the filters cannot be made arbitrarily narrow without introducing unacceptable gaps in the frequency coverage.

A naïve conclusion would thus be that the number of electrodes is the limiting factor. If this was true, then we would expect a 22 channel implant to have performance much better than a 12 channel implant. The evidence suggests otherwise; for example Friesen, Shannon et al. (2001) found that the speech perception of Nucleus 22 recipients improved as the number of channels used increased from four to twelve, but there was no significant improvement with more than twelve channels. Vandali, Sucher et al. (2005) found no difference in sung vowel pitch ranking scores between 10 channel CIS and 22 channel ACE. Using virtual channels to increase the number of stimulation channels beyond the number of physical electrodes appears to provide little benefit (§6.5.1) (Wilson et al. 1995). The issue is not that there are insufficient electrodes, but that there is inadequate spatial resolution in the stimulation patterns. Current steering does not improve spatial resolution (§6.5.2, Figure 6.10), and so is unlikely to provide any benefit.

It was argued in the previous section that a strong pitch sensation requires a specific phase relationship between the nerve firing times across a local region of the cochlea. Thus both fine spatial resolution and fine temporal resolution are required. It is simply not possible to reproduce this excitation pattern with a cochlear implant today. An electrical stimulation pulse causes neurons within a relatively broad region of the cochlea to fire at essentially the same time.

If the normal excitation pattern cannot be achieved, what can be done to improve cochlear implant pitch perception?

This thesis has shown that place pitch can convey a melody, and it could be a useful cue for fundamental frequencies above 500 Hz. However, it is hard to see how it can be utilised in the voice pitch frequency range. Electrode place is required to convey the overall spectral shape of the sound; in particular, to convey formants in speech.

Cochlear implant recipients are presently forced to rely on the relatively weak pitch produced by purely temporal means. Strategies that attempt to provide within-channel temporal cues (Table 6.2), such as HWR and TPS, are hampered by the broad spread of excitation. A neuron receives stimulation from perhaps four or five neighbouring electrodes, so that the temporal
cues presented on each channel are smeared together. Conversely, strategies that apply a global
temporal cue, such as F0 Modulation and MEM, are more resilient. Because the modulation on
all channels is in-phase, the temporal cue persists despite the spread of excitation. The penalty is
that only one pitch at a time can be presented, thus the full richness of music played by multiple
instruments cannot be conveyed. There also seems little prospect for using pitch differences to
segregate one voice from a competing speaker (§ 2.6). A further limitation is the restricted
frequency range for temporal cues of about two octaves, from 75 to 300 Hz (§ 3.7). Even within
this range, the difference limens are often many semitones (§ 9.2). Despite these limitations,
there may be a benefit for tonal languages.

The best means of improving pitch perception with present cochlear implant technology is to
make use of residual hearing (§ 6.6, § 11.5). Although the spatio-temporal response to a
harmonic tone is degraded in a hearing-impaired cochlea (Moore and Carlyon 2005), it appears
superior to that obtained with a cochlear implant.

Looking to the future, the obvious goal is to reproduce the excitation pattern of the normal ear.
Present models of the response to acoustic stimulation are sufficiently accurate for this purpose
(Figure 14.2a). Attempts to utilise such models in cochlear implants have so far been based on
simply sampling the acoustic model output to determine the electrical stimuli (§ 6.2.5). The flaw
in this approach is that it fails to account for the spread of electrical excitation. The solution is to
develop a model of electrical stimulation (Figure 14.2b), which predicts the neural excitation for
any specified electrical stimuli. If this model can be inverted, it can be used in the sound
processor (Figure 14.2c) to determine the electrical stimuli that will produce the desired neural
excitation pattern. This obviously requires better spatial and temporal control of stimulation
than is achieved in today’s implants.

Figure 14.2 Using models in cochlear implants
(a) Acoustic stimulation model (b) Electrical stimulation model
(c) Cochlear implant sound processor
The first steps along this path have been taken by van den Honert (van den Honert and Kelsall 2007). Firstly, an electrical stimulation model (Figure 14.2b) is created for each recipient. Each electrode is stimulated in turn in monopolar mode while measuring the voltage at every remaining electrode. This gives a transimpedance matrix. Inverting the transimpedance matrix gives the inverse model needed in Figure 14.2c. Initial psychophysical experiments have demonstrated improved spatial resolution (van den Honert and Kelsall 2007). A great deal of work remains to explore this promising new area.
15 Appendix 1: Subjects

The subject details are listed in Table 15.1. The first letter of the subject identifier indicates the subject location, either Sydney or Melbourne. The subjects were post-lingually deafened adults with the Nucleus 24 (CI24M, CI24R) or Nucleus Freedom (CI24RE) cochlear implant system. The last column indicates the per-channel stimulation rate used in their usual processor. Most subjects used the ACE strategy; a rate of 250 pps indicates SPEAK.

<table>
<thead>
<tr>
<th>Subject ID</th>
<th>Sex</th>
<th>Age</th>
<th>Aetiology</th>
<th>Implant use (years)</th>
<th>Implant Type</th>
<th>Electrode Type*</th>
<th>Rate (pps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>S01</td>
<td>F</td>
<td>43</td>
<td>Progressive</td>
<td>1.4</td>
<td>CI24R</td>
<td>CS</td>
<td>900</td>
</tr>
<tr>
<td>S02</td>
<td>M</td>
<td>78</td>
<td>Progressive</td>
<td>4.1</td>
<td>CI24M</td>
<td>St</td>
<td>250</td>
</tr>
<tr>
<td>S03</td>
<td>F</td>
<td>56</td>
<td>Meniere's</td>
<td>3.7</td>
<td>CI24R</td>
<td>St</td>
<td>900</td>
</tr>
<tr>
<td>S04</td>
<td>F</td>
<td>50</td>
<td>Unknown</td>
<td>4.2</td>
<td>CI24M</td>
<td>St</td>
<td>900</td>
</tr>
<tr>
<td>S05</td>
<td>M</td>
<td>54</td>
<td>Meniere's</td>
<td>2.6</td>
<td>CI24M</td>
<td>St</td>
<td>720</td>
</tr>
<tr>
<td>S06</td>
<td>M</td>
<td>37</td>
<td>Mumps</td>
<td>5.0</td>
<td>CI24R</td>
<td>CS</td>
<td>1800</td>
</tr>
<tr>
<td>S07</td>
<td>M</td>
<td>26</td>
<td>Unknown</td>
<td>0.9</td>
<td>CI24R</td>
<td>St</td>
<td>720</td>
</tr>
<tr>
<td>S08</td>
<td>F</td>
<td>86</td>
<td>Unknown</td>
<td>1.5</td>
<td>CI24R</td>
<td>CA</td>
<td>900</td>
</tr>
<tr>
<td>S09</td>
<td>M</td>
<td>72</td>
<td>Unknown</td>
<td>1.2</td>
<td>CI24RE</td>
<td>CA</td>
<td>1800</td>
</tr>
<tr>
<td>S10</td>
<td>M</td>
<td>88</td>
<td>Unknown</td>
<td>3.3</td>
<td>CI24R</td>
<td>St</td>
<td>900</td>
</tr>
<tr>
<td>S11</td>
<td>F</td>
<td>46</td>
<td>Meningococcal</td>
<td>2.2</td>
<td>CI24RE</td>
<td>St</td>
<td>900</td>
</tr>
<tr>
<td>M01</td>
<td>F</td>
<td>64</td>
<td>Unknown</td>
<td>3.2</td>
<td>CI24R</td>
<td>CS</td>
<td>900</td>
</tr>
<tr>
<td>M02</td>
<td>M</td>
<td>65</td>
<td>Noise induced</td>
<td>1.9</td>
<td>CI24R</td>
<td>CA</td>
<td>720</td>
</tr>
<tr>
<td>M03</td>
<td>M</td>
<td>78</td>
<td>Unknown</td>
<td>4.5</td>
<td>CI24R</td>
<td>CS</td>
<td>900</td>
</tr>
<tr>
<td>M04</td>
<td>F</td>
<td>84</td>
<td>Unknown</td>
<td>5.1</td>
<td>CI24R</td>
<td>CS</td>
<td>250</td>
</tr>
<tr>
<td>M05</td>
<td>F</td>
<td>78</td>
<td>Unknown</td>
<td>5.6</td>
<td>CI24R</td>
<td>CS</td>
<td>250</td>
</tr>
<tr>
<td>M06</td>
<td>F</td>
<td>74</td>
<td>Unknown</td>
<td>3.2</td>
<td>CI24R</td>
<td>CA</td>
<td>1200</td>
</tr>
<tr>
<td>M07</td>
<td>M</td>
<td>67</td>
<td>Meniere's</td>
<td>5.5</td>
<td>CI24R</td>
<td>CS</td>
<td>250</td>
</tr>
<tr>
<td>M08</td>
<td>F</td>
<td>41</td>
<td>Unknown</td>
<td>2.0</td>
<td>CI24R</td>
<td>CA</td>
<td>250</td>
</tr>
<tr>
<td>M09</td>
<td>F</td>
<td>73</td>
<td>Familial</td>
<td>5.0</td>
<td>CI24R</td>
<td>CS</td>
<td>900</td>
</tr>
<tr>
<td>M10</td>
<td>F</td>
<td>55</td>
<td>Unknown</td>
<td>1.0</td>
<td>CI24R</td>
<td>CA</td>
<td>500</td>
</tr>
<tr>
<td>M11</td>
<td>F</td>
<td>61</td>
<td>Unknown</td>
<td>6.1</td>
<td>CI24R</td>
<td>CS</td>
<td>900</td>
</tr>
<tr>
<td>M12</td>
<td>M</td>
<td>69</td>
<td>Familial</td>
<td>4.3</td>
<td>CI24R</td>
<td>CS</td>
<td>900</td>
</tr>
<tr>
<td>M13</td>
<td>M</td>
<td>75</td>
<td>Unknown</td>
<td>6.9</td>
<td>CI24R</td>
<td>CS</td>
<td>900</td>
</tr>
<tr>
<td>M14</td>
<td>M</td>
<td>53</td>
<td>Unknown</td>
<td>9.9</td>
<td>CI24M</td>
<td>St</td>
<td>250</td>
</tr>
<tr>
<td>M15</td>
<td>F</td>
<td>58</td>
<td>Familial</td>
<td>7.2</td>
<td>CI24R</td>
<td>CS</td>
<td>1200</td>
</tr>
<tr>
<td>M16</td>
<td>M</td>
<td>76</td>
<td>Meniere's</td>
<td>5.5</td>
<td>CI24R</td>
<td>CS</td>
<td>900</td>
</tr>
</tbody>
</table>

Table 15.1. Subject details

* Electrode type: St: Straight; CS: Contour; CA: Contour Advance
Table 15.2 is a cross-reference listing the experiments participated in by each of the Sydney subjects. Melbourne subjects M01 – M10 participated only in the Modified Melodies bimodal study (§11.5). Melbourne subjects M11 – M16 participated only in the Freedom HWR study (§12.3.1).

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Section</th>
<th>S01</th>
<th>S02</th>
<th>S03</th>
<th>S04</th>
<th>S05</th>
<th>S06</th>
<th>S07</th>
<th>S08</th>
<th>S09</th>
<th>S10</th>
<th>S11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pitch rank electrodes</td>
<td>8.2</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pitch rank dual-electrode</td>
<td>8.3</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pitch rank pure tone ACE</td>
<td>8.4</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pitch rank SPP MPP</td>
<td>9.2</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pitch match SPP MPP</td>
<td>9.3</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Warble discrimination</td>
<td>9.4</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pitch rank pure tone HWG TPS</td>
<td>10.2</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pitch rank sung vowel</td>
<td>10.3</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Modified Melodies harmonic tone ACE</td>
<td>11.4</td>
<td>*</td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Modified Melodies bimodal</td>
<td>11.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Freedom ACE HWR</td>
<td>12</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Modified Melodies pure tone ACE</td>
<td>13</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 15.2. Sydney subject experiment participation
Appendix 2: Statistics

16 Appendix 2: Statistics

Most of the experiments in this thesis use forced choice procedures, and so the results can be analysed with the binomial probability distribution. This chapter describes the hypothesis-testing approach used in this thesis.

A binomial experiment consists of repeated trials, where the outcome of each trial is labelled either success or failure, and where the probability of success remains constant from trial to trial. (Equivalently we may use the less pejorative labels correct or incorrect.) If the probability of success on a single trial is \( p \), then the probability of obtaining \( x \) successes out of \( n \) trials is:

\[
b(x, n, p) = C(n, x) \ p^x (1 - p)^{n-x}
\]

where \( C(n,x) \) is the number of combinations of \( n \) objects taken \( x \) at a time. Equation 16.1 is referred to as the binomial probability density function. The variance is:

\[
\sigma^2 = n \ p (1 - p)
\]

Traditionally, the analysis of experimental results relied upon the statistical tables found as appendices in every statistics textbook. The speed of today's personal computers makes a more direct approach feasible. It will be illustrated in the following sections using the MATLAB Statistics toolbox. Excerpts from a MATLAB session will be shown indented, in Courier font. Lines that begin with the command prompt ">>" indicate those typed by the user, and the following lines show the response. The execution times quoted in the following sections were obtained on a 1.73 GHz PC.

16.1 Evaluating a single binomial experiment

In evaluating the result of a forced-choice experiment, the first question is whether the subject was merely guessing. This straightforward case will be dealt with in some detail, so that the reasoning behind the more complicated cases is clearer. For a two-alternative forced-choice (2AFC) procedure, we consider the null hypothesis that the probability of success on each trial was \( p = 0.5 \). The typical textbook approach is to approximate the discrete binomial distribution with a continuous normal distribution, and then calculate a Z-score. The normal approximation is reasonably accurate if both \( x \) and \( n - x \) are greater than five.

In most experiments in this thesis, trials were done in blocks of eight. Thus if only one or two blocks were performed, the normal approximation will be inaccurate, especially if scores at or near 100% were obtained. The number of cochlear implant recipients who are willing and available to act as unpaid volunteers in experiments is limited, and we should not waste their
time. The subject should be asked to perform the least number of trials that are sufficient to
demonstrate an effect. We should not do more trials merely for the convenience of using the
normal approximation. Instead, we should apply statistical methods that cope with relatively
small numbers of trials. If an effect is so small that more than 32 trials are required to show
statistical significance, then the effect is of little practical importance.

In place of the normal approximation, we can apply the binomial distribution directly. The
number of possible outcomes is $n + 1$, i.e. $x = 0$ up to $n$. The binomial distribution (Equation
16.1) tells us the probability of occurrence of each outcome. The MATLAB function to
calculate the binomial probability density function is:

```matlab
>> binopdf(x, n, p)
```

where $x$, $n$ and $p$ have the same meaning as in Equation 16.1. The argument $x$ is permitted to be
a vector. If the subject was guessing, then for one block of 8 trials, the probability of each of the
9 possible outcomes (0 to 8 correct) is:

```matlab
>> b = binopdf(0:8, 8, 0.5)
b =
    0.0039 0.0312 0.1094 0.2187 0.2734 0.2187 0.1094 0.0312 0.0039
```

For example, say a subject obtained a score of 7 correct out of 8 trials. What should we
conclude? The probability distribution is plotted as a bar graph in Figure 16.1. The scores equal
to or higher than the observed score (sometimes called the extreme outcomes, or the tail of the
distribution) are shown in black. The probability of getting a score of 7 or more by guessing is:

```matlab
>> p = sum(binopdf(7:8, 8, 0.5))
p = 0.0352
```

Since this probability is less than the criterion value $\alpha = 0.05$ that is generally accepted for
statistical significance, we conclude that it is unlikely that the null hypothesis is true, i.e. the
subject was most likely using some cue in the stimuli to obtain a good score. A one-sided test
was used, which is appropriate if there was no expectation that the score could have been worse
than chance.
16.2 Comparing two binomial experiments

In many experiments, we are interested in comparing the subject's performance under two different conditions (e.g. comparing pitch-ranking scores with two different signal processing strategies). We would like to determine whether one condition gives better performance than another. Consider an experiment where, after 8 trials in each condition, the subject scored 3 correct with condition A, and 7 correct with condition B. These results can be set out as shown in Table 16.1, known as a contingency table. Intuitively, it appears that the subject was guessing in condition A, but was making use of a definite cue in condition B. How confident should we be in this conclusion, especially since we have such a small number of trials? We need to know how likely it is that the observed difference in scores is solely due to random variation. There are several ways of answering this question.

<table>
<thead>
<tr>
<th></th>
<th>Number of successes</th>
<th>Number of failures</th>
<th>Number of trials</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condition A</td>
<td>3</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>Condition B</td>
<td>7</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>Totals</td>
<td>10</td>
<td>6</td>
<td>16</td>
</tr>
</tbody>
</table>

Table 16.1. Contingency table example
### Table 16.2. Contingency table notation

<table>
<thead>
<tr>
<th></th>
<th>Number of successes</th>
<th>Number of failures</th>
<th>Number of trials</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Condition A</strong>&lt;br&gt;(probability (p_a))</td>
<td>(x_a)</td>
<td>(y_a)</td>
<td>(n_a = x_a + y_a)</td>
</tr>
<tr>
<td><strong>Condition B</strong>&lt;br&gt;(probability (p_b))</td>
<td>(x_b)</td>
<td>(y_b)</td>
<td>(n_b = x_b + y_b)</td>
</tr>
<tr>
<td><strong>Totals</strong></td>
<td>(x_t = x_a + x_b)</td>
<td>(y_t = y_a + y_b)</td>
<td>(n_t = n_a + n_b)</td>
</tr>
</tbody>
</table>

The notation that will be used is shown in Table 16.2. The appropriate null hypothesis is that both conditions give the same performance, i.e. both conditions have the same probability of success, which we denote \(p_0\):

\[
H_0: p_a = p_b = p_0
\]  

(16.3)

If the null hypothesis is true, then instead of two experiments (one for condition A, and one for condition B) there was effectively one "pooled" binomial experiment, which yielded \(x_t\) correct out \(n_t\) trials (the bottom row of Table 16.2), thus the best estimate of the probability \(p_0\) is:

\[
p_0 = \frac{x_t}{n_t} = \frac{x_a + x_b}{n_a + n_b}
\]  

(16.4)

#### 16.2.1 Z-test

Most textbooks describe a Z-test. If the null hypothesis is true, the following variable has approximately a standard normal distribution:

\[
z = \frac{p_b - p_a}{\sqrt{p_0(1-p_0)(1/n_a + 1/n_b)}}
\]  

(16.5)

This method was implemented by the author as a MATLAB function, which can be called as:

\[
\text{>> Difference\_between\_proportions([3,7], 8, 'z')}
\]

The last argument is a string specifying the statistical method to be used (alternative methods will be described in the following sections). The one-sided Z-test gives the result \(p = 0.019\), which is statistically significant. However, as noted earlier, the normal approximation requires both \(x\) and \(y\) to be greater than five, so it may be seriously inaccurate here.

#### 16.2.2 Monte Carlo simulation

Simon (1997) advocated the unorthodox approach of simulating the experiment on a computer,
sometimes known as the Monte Carlo method. The MATLAB command to generate pseudo-random numbers with a binomial distribution is:

```matlab
>> sim_x = binornd(n, p, r, c)
```

This produces a matrix `sim_x` (having `r` rows and `c` columns) of random integers in the range 0 to `n`, according to a binomial distribution, where `n` is the number of trials and `p` is the probability of success in each trial. To simulate the experiment in Table 16.1, we firstly calculate the null-hypothesis probability `p0` (according to Equation 16.4):

```matlab
>> p0 = (3 + 7)/(8 + 8)
p0 = 0.6250
```

Then we simulate 100,000 runs of the experiment:

```matlab
>> sim_xa = binornd(8, p0, 1, 100000);
>> sim_xb = binornd(8, p0, 1, 100000);
```

The row-vector `sim_xa` contains the score (the number of successes) for condition A for each run, and the row-vector `sim_xb` contains the score for condition B for each run. Executing this took less than 0.2 seconds, and the results of the first 10 simulated runs were:

```matlab
>> sim_xa(1:10)
ans =
5 3 5 3 4 7 4 2 5 7
>> sim_xb(1:10)
ans =
4 4 5 5 6 6 6 4 6 4
```

Next we find the difference between the corresponding pairs of scores, and display the first 10 of them:

```matlab
>> sim_d = sim_xb - sim_xa;
>> sim_d(1:10)
ans =
-1 1 0 2 2 -1 2 2 1 -3
```

The observed difference in scores (Table 16.1) was \(x_b - x_a = 7 - 3 = 4\). Out of the 100,000 simulated runs, the number of runs that had a difference in scores of 4 or more was:

```matlab
>> num_extreme = sum(sim_d >= 4)
num_extreme = 3274
```

Thus, according to the simulation, a difference in scores of 4 or more is a relatively rare event. Finally we divide by the number of simulations to yield an estimate of the probability that the observed difference in scores (or an even larger difference) would occur by chance, if condition A and B actually gave the same performance:

```matlab
>> p_extreme = num_extreme/100000
p_extreme =
``
This is statistically significant \((p < 0.05)\); therefore, we reject the null hypothesis, and conclude that condition B gave better performance than condition A. In this case, a one-sided test was used, which is appropriate in the situation where (for example) condition A is the "standard" processing strategy, and condition B is a new processing strategy intended to provide better performance.

The Monte Carlo method was implemented by the author in the function `Difference_between_proportions`, and can be invoked as:

```python
>> Difference_between_proportions([3,7], 8, 'monte-carlo')
```

Although this example has the same number of trials for both conditions, it is straightforward to handle cases where the number of trials differs: the difference between the proportions correct is calculated, rather than the difference in the number of successes.

An important issue is how to choose the number of runs to simulate. Each simulated run can itself be considered a single trial of a binomial experiment with the probability of success given by \( p_{\text{extreme}} \). Inserting our estimate of \( p_{\text{extreme}} \) into Equation 16.2, the standard deviation of the number of successes (i.e. the number of extreme outcomes) in 100,000 trials is estimated as:

```python
>> std_dev = sqrt(100000 * p_extreme * (1-p_extreme))
std_dev =
56.4898
```

which is small compared to the 3274 extreme outcomes seen in the simulation. It is very likely that the simulated estimate is within two standard deviations of the "true" number of extreme outcomes. Because the number of simulation runs is so large, the normal approximation is valid, so an approximate 95% confidence interval for the probability is:

```python
>> p_delta = norminv(0.975) * std_dev / 100000
p_delta =
0.0011
>> confidence_interval = [p_extreme - p_delta, p_extreme + p_delta]
confidence_interval =
0.0319  0.0341
```

As a further check, repeating the entire simulation 10 times gave the probability estimates:

```
0.0329  0.0324  0.0334  0.0328  0.0331  0.0339  0.0343  0.0341
0.0333  0.0340
```

The mean was 0.0334. The observed variation is in line with the estimated precision, and is certainly sufficient to justify the decision to reject the null hypothesis.
16.2.3 Evaluation of entire PDF

For those who are uncomfortable with the idea of simulating the experiment, an alternative approach is outlined below. Although simple and intuitive, it does not appear to be presented in the literature, probably because the computing power required has only recently become available.

To compare the results of two binomial experiments (condition A and condition B), we again consider the null hypothesis that the two conditions have the same probability of success (Equation 16.3) and calculate the pooled estimate $p_0$ according to Equation 16.4. For condition A, the number of possible outcomes is $n_a + 1$, i.e. $x_a = 0$ up to $n_a$. For the example in Table 16.1, the probability of each possible outcome is:

```matlab
>> pdf_a = binopdf(0:8, 8, p0)
pdf_a =
    0.0004 0.0052 0.0304 0.1014 0.2112 0.2816 0.2347 0.1118 0.0233
```

Similarly, we can calculate the probability of each possible outcome for condition B (which is the same as for condition A in this example, because the number of trials is the same). The number of possible pairs of outcomes for the two binomial experiments is simply $(n_a + 1)$ multiplied by $(n_b + 1)$. As the two experiments are independent, the probability of each one of these pairs is the product of the two individual outcomes. In MATLAB, we can transpose `pdf_a` to give a column vector, and then multiply it by the row vector `pdf_b`, to produce a matrix `pdf_ab`, with $(n_a + 1)$ rows and $(n_b + 1)$ columns:

```matlab
>> pdf_ab = pdf_a' * pdf_b
pdf_ab =
    0.0000 0.0000 0.0000 0.0000 0.0001 0.0001 0.0001 0.0000 0.0000
    0.0000 0.0000 0.0002 0.0005 0.0011 0.0015 0.0012 0.0006 0.0001
    0.0000 0.0002 0.0009 0.0031 0.0064 0.0086 0.0071 0.0034 0.0007
    0.0000 0.0005 0.0031 0.0103 0.0214 0.0286 0.0238 0.0113 0.0024
    0.0001 0.0011 0.0064 0.0214 0.0446 0.0595 0.0496 0.0236 0.0049
    0.0001 0.0015 0.0086 0.0286 0.0595 0.0793 0.0661 0.0315 0.0066
    0.0001 0.0012 0.0071 0.0238 0.0496 0.0661 0.0551 0.0262 0.0055
    0.0000 0.0006 0.0034 0.0113 0.0236 0.0315 0.0262 0.0125 0.0026
    0.0000 0.0001 0.0007 0.0024 0.0049 0.0066 0.0055 0.0026 0.0005
```

The element at row $k_a$ and column $k_b$ is the joint probability of getting $k_a$ successes out of $n_a$ trials in condition A, and $k_b$ successes out of $n_b$ trials in condition B. As a check on MATLAB's accuracy, the sum of these entries should be 1.0:

```matlab
>> 1 - sum(pdf_ab(:))
an =
    4.2188e-015
```

This two-dimensional joint probability density function is shown as an image in Figure 16.2. Probabilities near zero are displayed as white, and the maximum probability in this function is...
displayed as black. The observed result \((x_a, x_b) = (3, 7)\) is marked with a small square. Keeping in mind that this probability density function represents the null hypothesis (condition A and condition B both having probability of \(p_0 = 0.6250\)), we can see that if the null hypothesis is true, then the observed result had a low probability.

![Figure 16.2 Joint probability density function for the null hypothesis](image)

To decide whether to reject the null hypothesis, we sum together the probabilities of all outcomes that have a difference greater than or equal to the observed difference, i.e. \(k_b - k_a \geq 4\). These extreme outcomes reside in the upper right corner of Figure 16.3 and are marked with small crosses. These calculations were also implemented in the author's MATLAB function, with the result:

```matlab
>> s = Difference_between_proportions([3,7], 8, 'brute-force');
>> s.p_extreme
ans =
   0.0336
```

Thus the null hypothesis is rejected. This exact result demonstrates the accuracy of the Monte Carlo simulation result, and the inaccuracy of the Z-test \((p = 0.019\) in §16.2.1). The `Difference_between_proportions` function also handles cases where the number of trials differs. The probability density matrix `pdf_ab` becomes rectangular, rather than square, and the difference between the proportions correct is calculated.
16.2.4 Fisher's exact test

For small numbers of trials, textbooks often recommend Fisher's exact test. We start with an observed contingency table in the format of Table 16.2. The reasoning then proceeds as follows: if the null hypothesis is true (condition A and condition B have equal probability), then we obtained \( x_t \) successes and \( y_t \) failures in the pooled experiment. How many ways could this set of successes and failures be partitioned between condition A and condition B? We list all possible contingency tables that have the same numbers of trials \( (n_a, n_b) \), and the same sub-totals \( (x_t, y_t) \) as the observed outcome. The number of such tables will be one more than the smallest of the sub-totals \( (x_t, y_t) \) in the observed table. In the example in Table 16.1, the smallest sub-total is \( y_t = 6 \), so we list the seven contingency tables shown in Table 16.3 (of which the second table listed is the observed outcome). The final column is explained below. Note that these seven tables all give the same pooled probability estimate, \( p_0 \) (Equation 16.4).

The next step is to determine the probability of occurrence for each of the listed tables. The probability of event F, given that event G is known to have occurred (the conditional probability of F given G) is:

\[
p(F \mid G) = \frac{P(F \cap G)}{P(G)}
\]  

(16.6)
<table>
<thead>
<tr>
<th></th>
<th>Number of successes</th>
<th>Number of failures</th>
<th>Number of trials</th>
<th>Conditional probability</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Condition A</strong></td>
<td>2</td>
<td>6</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td><strong>Condition B</strong></td>
<td>8</td>
<td>0</td>
<td>8</td>
<td>0.0035</td>
</tr>
<tr>
<td><strong>Totals</strong></td>
<td>10</td>
<td>6</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td><strong>Condition A</strong></td>
<td>3</td>
<td>5</td>
<td>8</td>
<td>0.0559</td>
</tr>
<tr>
<td><strong>Condition B</strong></td>
<td>7</td>
<td>1</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td><strong>Totals</strong></td>
<td>10</td>
<td>6</td>
<td>16</td>
<td>0.2448</td>
</tr>
<tr>
<td><strong>Condition A</strong></td>
<td>5</td>
<td>3</td>
<td>8</td>
<td>0.3916</td>
</tr>
<tr>
<td><strong>Condition B</strong></td>
<td>5</td>
<td>3</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td><strong>Totals</strong></td>
<td>10</td>
<td>6</td>
<td>16</td>
<td>0.3916</td>
</tr>
<tr>
<td><strong>Condition A</strong></td>
<td>6</td>
<td>2</td>
<td>8</td>
<td>0.2448</td>
</tr>
<tr>
<td><strong>Condition B</strong></td>
<td>4</td>
<td>4</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td><strong>Totals</strong></td>
<td>10</td>
<td>6</td>
<td>16</td>
<td>0.2448</td>
</tr>
<tr>
<td><strong>Condition A</strong></td>
<td>9</td>
<td>1</td>
<td>8</td>
<td>0.0559</td>
</tr>
<tr>
<td><strong>Condition B</strong></td>
<td>7</td>
<td>3</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td><strong>Totals</strong></td>
<td>10</td>
<td>6</td>
<td>16</td>
<td>0.0559</td>
</tr>
<tr>
<td><strong>Condition A</strong></td>
<td>8</td>
<td>0</td>
<td>8</td>
<td>0.0035</td>
</tr>
<tr>
<td><strong>Condition B</strong></td>
<td>2</td>
<td>6</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td><strong>Totals</strong></td>
<td>10</td>
<td>6</td>
<td>16</td>
<td>0.0035</td>
</tr>
</tbody>
</table>

Table 16.3. Fisher-Irwin contingency tables example

Thus the probability of obtaining the pair of scores \((x_a, x_b)\), given that the sub-totals are \((x_t, y_t)\) is:

\[
P(x_a, x_b) = \frac{b(x_a, n_a, p_0) \cdot b(x_b, n_b, p_0)}{b(x_t, n_t, p_0)}
= \frac{C(n_a, x_a) \cdot p_0^{x_a} (1-p_0)^{n_a-x_a} \cdot C(n_b, x_b) \cdot p_0^{x_b} (1-p_0)^{n_b-x_b}}{C(n_t, x_t) \cdot p_0^{x_t} (1-p_0)^{n_t-x_t}}
= \frac{C(n_a, x_a) \cdot C(n_b, x_b)}{C(n_t, x_t)}
= \frac{n_a! \cdot n_b! \cdot x_a! \cdot y_a!}{n_t! \cdot x_a! \cdot y_a! \cdot x_b! \cdot y_b!}
\]

(16.7)

Note the cancellation of all the terms involving \(p_0\) to obtain the third line of Equation 16.7. Continuing the example, the conditional probabilities are listed in the final column of Table 16.3. As a check on calculations, their sum should be 1.0. Finally we sum the conditional probabilities of the extreme outcomes, i.e. those tables where the difference between condition A and B was greater than or equal to the observed difference of 4, i.e. the first two sub-tables in Table 16.3. For completeness, these calculations were also implemented in the author's
MATLAB function, with the result:

```matlab
>> s = Difference_between_proportions([3,7], 8, 'fisher);
>> s.p_extreme
ans =
    0.0594
```

Unhappily, this does not agree with the result obtained by the direct evaluation of the joint probability density function, or the simulation result. The discrepancy is caused by the fact that Fisher's exact test does not consider all possible outcomes of the two binomial experiments. This is illustrated in Figure 16.4, which once again shows an image plot of the two-dimensional joint probability density function for the example of Table 16.1, with the observed outcome marked with a small square, and the seven outcomes that are considered in Fisher's exact test marked with small circles. By only considering outcomes that have the same sub-totals \((x_i, y_i)\) as the observed outcome, an erroneous dependency is introduced between conditions A and B. For the example (Table 16.3), if condition A scored 4 out of 8, then the only outcome considered for condition B is 6 out of 8. This does not reflect the actual experiment, where conditions A and B are independent. Effectively, Fisher's exact test evaluates the probabilities along a diagonal slice through the probability density function, reducing the two-dimensional problem to a one-dimensional one. It provides a crude approximation to the true result because the slice is characteristic of the shape of the entire function, and its direction is orthogonal to the edge defining the extreme region (see Figure 16.3). Its appeal also lies in the simple form of Equation 16.7, which allows hand calculation. However, as it is now eminently practical to evaluate the entire two-dimensional probability density function, Fisher's exact test should be regarded as a relic of a bygone era.

![Figure 16.4 Joint probability density function showing the outcomes considered in Fisher's exact test](image-url)
16.3 Dependence of performance on a parameter

The previous section compared the results of two binomial experiments. This section extends the analysis to handle more than two conditions. The observed result is a vector \([x_1, x_2, \ldots, x_M]\), where the number of conditions is denoted \(M\), and \(x_m\) is the number of successes in the \(m\)th condition. Usually each condition has the same number of trials, \(n\). The different conditions often correspond to different values of some parameter of the stimulus. In many experiments in this thesis, we are interested in how the subject's perception of pitch varies with frequency. In this case, each condition corresponds to a different base frequency. Given a limit on the total number of trials the subject can perform, there is clearly a trade-off between the number of different conditions and the number of trials per condition. To ascertain the relation between frequency and the subject's performance, we would prefer to measure performance at many different frequencies, which implies a relatively small number of trials per condition.

When examining the results, the first question is whether the overall performance across all conditions (e.g. across the entire frequency range) was better than chance. Once again, we consider the pooled experiment, for which the overall proportion of successes is:

\[
P_0 = \frac{\sum x_m}{nM} = \frac{x_t}{n_t}
\]  

(16.8)

We test the null hypothesis that \(p = 0.5\) (as described in §16.1). This test is sensitive because the total number of trials \((nM)\) is relatively large.

We can also ask whether the performance for any one condition (e.g. at one frequency) was better than chance. This can also be tested as described in §16.1. However, if the subject was tested under many conditions, all of which had true performance at chance level, then we would expect that occasionally (about 5% of the time) the subject would obtain a score that appeared to be significantly greater than chance. Hence, we will not attach much importance to an individual result: we are more interested in the overall trend. This leads to the third question, which is whether the performance differs amongst the different conditions (e.g. does the performance depend on frequency?). If performance does vary, then it is worthwhile seeking an explanation, and perhaps fitting a perceptual model to the results. Therefore, we need to know how likely it is that an apparent pattern in the scores is simply due to random fluctuation.

As a concrete example, one result from an experiment that was described in §8.4 will be analysed. The number of trials correct out of \(n = 16\) trials for subject S05, for pitch-ranking pure tones with six-semitone spacing, for four conditions (base frequencies of 99, 140, 198 and 281 Hz) were (in MATLAB syntax):

\[
>> x = [8; 12; 11; 16];
\]
The scores have a large spread, from chance to a perfect score. Is a spread as large as this likely to occur by chance? We consider the null hypothesis that the probability of success in a trial is actually constant (but not necessarily 50%) across this frequency range. As before, the best estimate of this constant probability of success is the overall proportion correct (i.e. the pooled estimate):

\[
p_0 = \frac{\text{mean}(x)}{16}
\]

\[
p_0 = 0.7344
\]

i.e. an overall score of 73% correct. When there were only two conditions, the difference between the two scores was an appropriate statistic (§16.2). For more than two conditions, the sample variance provides a good measure of the spread. For a set of \(M\) scores \((x_1, x_2, \ldots, x_M)\), the sample variance is:

\[
s^2 = \frac{\sum (x_m - \bar{x})^2}{M - 1}
\]

where \(\bar{x}\) is the sample mean. The sample variance \(s^2\) is an unbiased estimator of the distribution variance \(\sigma^2\). Continuing the example of pure-tone pitch-ranking scores for subject S05, if the null hypothesis is true, the binomial distribution variance is given by Equation 16.2, with \(n = 16\) and \(p = p_0 = 73\%\), i.e. in MATLAB:

\[
\text{>> v0 = 16 * p0 * (1 - p0)}
\]

\[
v0 = 3.1211
\]

In comparison, the subject's results have a sample variance of:

\[
\text{>> v = var(x)}
\]

\[
v = 10.9167
\]

We see that the sample variance is much larger than would be expected if performance were independent of frequency. To quantify the significance of this result, we again turn to the Monte Carlo method.

16.3.1 Monte Carlo simulation

We simulate 100,000 runs of the experiment:

\[
\text{>> sim_x = binornd(16, p0, 4, 100000)};
\]
Appendix 2: Statistics

The matrix \( \text{sim}_x \) has 4 rows and 100,000 columns. Each column represents the number of successes of the four conditions for one run of the experiment. Next, we calculate the sample variance for each run, and find the proportion of runs that had a sample variance equal to or larger than the observed sample variance:

\[
\begin{align*}
\text{>> sim}_v & = \text{var}(\text{sim}_x); \\
\text{>> p\_extreme} & = \text{sum}(\text{sim}_v \geq v) / 100000 \\
\text{p\_extreme} & = 0.0173
\end{align*}
\]

This is statistically significant \( (p < 0.05) \); therefore, we reject the null hypothesis, and conclude that the performance does depend on frequency. The execution time was 0.7 seconds. This method was implemented by the author as a MATLAB function:

\[
\begin{align*}
\text{>> s} & = \text{Spread\_of\_proportions}(x, 16, 'monte\_carlo'); \\
\text{>> s.v\_sig} & \\
\text{ans} & = 0.0175
\end{align*}
\]

16.3.2 Evaluation of entire PDF

An earlier section (§16.2) presented an exact method of comparing two binomial scores, where every possible outcome was evaluated. For multiple conditions, a similar "brute force" exact method is feasible if the experiment is relatively small, so it was also implemented in the author's \text{Spread\_of\_proportions} function. In the example considered above, there were four conditions of 16 trials each. The number of possible outcomes is \( (n + 1)^M = 17^4 = 83521 \), which is manageable. The function iterates through every possible outcome. For each outcome, the sample variance is calculated. If that sample variance is greater than or equal to the observed sample variance, then the probability of that outcome is computed (the product of the four binomial probabilities), and added to a running sum of the cumulative probability. For the example scores of subject S05, this exact method gives:

\[
\begin{align*}
\text{>> s} & = \text{Spread\_of\_proportions}(x, 16, 'brute\_force'); \\
\text{>> s.v\_sig} & \\
\text{ans} & = 0.0170
\end{align*}
\]

It required 8 seconds (about ten times longer than the Monte Carlo simulation). The execution time would be prohibitive for a larger number of conditions.

16.3.3 Chi-square test

An alternative approach would be possible if \( x \) was normally distributed, because then the sampling distribution of the random variable:
\[ \chi^2 = \frac{(M - 1)s^2}{\sigma^2} \]  

(16.10)

would have a chi-square distribution with \( M - 1 = 3 \) degrees of freedom. The distribution of \( x \) is of course binomial, but if we assume that the normal distribution is a reasonable approximation, then we can scale the observed sample variance according to Equation 16.10:

```matlab
>> chi2 = v * (3/v0)
chi2 =
  10.4931
```

and test this variable against the chi-square cumulative density function. For completeness, this method was also implemented in the author's `Spread_of_proportions` function:

```matlab
>> s = Spread_of_proportions(x, 16, 'chi square');
>> s.v_sig
ans =
  0.0148
```

Compared to the exact result of 0.0170, the relative error is 13%. We can use the Monte Carlo simulation to show the disparity between the distribution of the scaled sample variance and the chi-square distribution. Firstly we scale the simulated sample variances according to Equation 16.10:

```matlab
>> sim_chi2 = sim_v * (3/v0);
```

A histogram of the scaled sample variance (``sim_chi2``) of the 100,000 simulated runs of the experiment is shown in Figure 16.5 (in black), together with a histogram of the chi-square distribution with 3 degrees of freedom (in red).

![Figure 16.5 Histogram of simulated scaled sample variance (black) and chi-square distribution (red)](image)
Perhaps surprisingly, the scaled sample variance does not have a smooth distribution. This is because the scores $x_m$ in Equation 16.9 take on discrete integer values clustered around the mean, and there are only $M = 4$ conditions. The chi-square distribution is plainly a crude approximation. The approximation would be much better if the number of trials per condition was more than 30.

### 16.3.4 Presentation of experimental results

The format shown in Table 16.4 (excerpted from Table 8.4) is used to report experimental results in the experimental chapters of the thesis. Each column contains one subject's results. The first row identifies the subject. The following rows contain the number correct for each condition. If an individual score is significantly better than chance ($p < 0.05$) it is marked with an asterisk, or with two asterisks if $p < 0.01$. The next row displays the number of trials per condition. The next row shows the pooled score (i.e. the overall score summed across all conditions) according to Equation 16.8, expressed as percent correct. The following row has a $p$-value indicating whether the overall score is significantly better than chance.

The last three rows indicate the amount of spread in the scores. The numerical value of the sample variance offers little insight. Instead, the table shows the sample standard deviation (the square root of the sample variance) of the scores, expressed as percent correct. In this example:

```matlab
>> pc = 100 * x / 16;
>> std_dev_pc = std(pc)
std_dev_pc =
    20.6502
```

which is rounded to 21% in Table 16.4. The next row shows the pooled estimate of the standard deviation if the null hypothesis was true (Equation 16.2), again expressed as the spread in percent correct:

```matlab
>> s0_pc = 100 * sqrt(v0)/16
s0_pc =
    11.0416
```

which is rounded to 11% in Table 16.4. The final row shows the $p$-value indicating whether the sample variance is significantly greater than the null hypothesis variance, i.e. the probability that the spread in the observed scores would occur by chance, if the performance were the same in all conditions. This $p$-value is calculated by the Monte Carlo simulation of sample variance, described above. In this example, the standard deviation is roughly twice as big as might be expected by chance, which is statistically significant ($p = 0.0173$, which is rounded to 0.02 in Table 16.4).
Occasionally the number of trials may differ across parameter values. In this case, the analysis uses the proportions correct instead of the number of successes. The results are reported in a similar format to Table 16.4, except that the score for each condition is shown as "$x / n\)$, i.e. the number of successes out of the number of trials (for example, Table 8.6 in §8.4).

The tabular format has the benefit of containing the complete experimental results, enabling a critical reader to verify the statistical analysis. In addition, a plot of the scores will accompany each table, to better visualise the trends in the data. As the statistics are given in the table, there is less need to clutter the plots with confidence intervals ("error bars").

### 16.4 Comparing two sets of binomial experiments

This section considers the case where there are two conditions, each of which was tested at multiple parameter values (with the same set of parameter values used for both conditions). We wish to compare the performance of the two conditions.

One result of an experiment to be described in §9.2 will be used as an example. The results for subject S02, for pitch-ranking two types of pulse sequence (denoted SPP and MPPU), across six frequency intervals, are shown in Table 16.5. We seek a statistical test of the hypothesis that SPP gives better performance than MPPU.
### Table 16.5 Subject S02 pitch ranking test results.

<table>
<thead>
<tr>
<th>Frequency Range</th>
<th>SPP</th>
<th>MPPU</th>
</tr>
</thead>
<tbody>
<tr>
<td>99 - 125 Hz</td>
<td>16**</td>
<td>24**</td>
</tr>
<tr>
<td>125 - 157 Hz</td>
<td>16**</td>
<td>22**</td>
</tr>
<tr>
<td>157 - 198 Hz</td>
<td>16**</td>
<td>24**</td>
</tr>
<tr>
<td>198 - 250 Hz</td>
<td>15**</td>
<td>23**</td>
</tr>
<tr>
<td>250 - 315 Hz</td>
<td>12*</td>
<td>13</td>
</tr>
<tr>
<td>315 - 397 Hz</td>
<td>10</td>
<td>11</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>SPP</th>
<th>MPPU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of trials</td>
<td>16</td>
<td>24</td>
</tr>
<tr>
<td>Pooled score</td>
<td>89%</td>
<td>81%</td>
</tr>
<tr>
<td>- Significance</td>
<td>&lt;0.00</td>
<td>&lt;0.00</td>
</tr>
<tr>
<td>Std dev of scores</td>
<td>16%</td>
<td>25%</td>
</tr>
<tr>
<td>- Pooled estimate</td>
<td>8%</td>
<td>8%</td>
</tr>
<tr>
<td>- Significance</td>
<td>&lt;0.00</td>
<td>&lt;0.00</td>
</tr>
</tbody>
</table>

#### 16.4.1 Pooled comparison

A simple approach is to pool (i.e. sum) the results for each condition, and compare them according to §16.2. The function `Difference_between_proportions` defaults to the "brute force" method, and has no difficulty handling this larger numbers of trials (96 and 144 trials):

```matlab
>> xa = [16; 16; 16; 15; 12; 10]; na = 16;
>> xb = [24; 22; 24; 23; 13; 11]; nb = 24;
>> s_pooled = Difference_between_proportions([sum(xa), sum(xb)], 6*[na, nb]);
>> s_pooled.p_extreme
ans =
   0.0680
```

Based upon the pooled scores, the two conditions are not significantly different. However, this method is flawed, because treating the pooled result for each condition as a single binomial experiment assumes that the probability of success is constant across the parameter set (i.e. independent of frequency). On the contrary, the standard deviation of scores, shown in Table 16.5, indicates that performance is very dependent on frequency.

#### 16.4.2 Repeated-measures t-test

Instead, we recognise that the experiment used a repeated-measures design. The large variation within each condition masks the difference between the two conditions. A paired comparison is a more sensitive test, because it eliminates the effect of the within-condition variability. As the
two conditions had different numbers of trials, we subtract the corresponding proportions correct:

```matlab
>> pa = xa / na;
>> pb = xb / nb;
>> pd = pa - pb
pd =
   0
   0.0833
   0
  -0.0208
   0.2083
  0.1667
```

The variable \(pd\) contains the difference in proportions correct at each frequency. At most frequencies, the SPP score was equal to or greater than the MPPU score. The standard statistical method to determine whether the observed difference in scores is significant is a t-test. We test the null hypothesis that the difference is less than or equal to zero (a one-sided test), using a significance level of \(\alpha = 0.05\):

```matlab
>> [h, p] = ttest(pd, 0, 0.05, 1)
```

\[h = 0\]

\[p = 0.0619\]

Thus, we fail to reject the null hypothesis (\(p > 0.05\)), i.e. according to the t-test, the two conditions are not significantly different. However, this is yet another case where the normal approximation is inaccurate (due to many scores being at or near 100%), so the t-test is not appropriate.

### 16.4.3 Monte Carlo simulation

Once more, we turn to the Monte Carlo method. We consider the null hypothesis that the two conditions give the same performance as each other at each frequency (although the performance may depend on frequency). If so, the best estimate of the probability of success at each frequency is:

```matlab
>> p0 = (xa + xb) / (na + nb)
p0 =
   1.0000
  0.9500
  1.0000
  0.9500
  0.6250
  0.5250
```

Note that the results are pooled across conditions (not across frequency), which is valid if the
null hypothesis is true. The mean difference between the proportions correct was:

```matlab
>> mpd = mean(pd)
mpd =
 0.0729
```

How often would a mean difference as large as this (or larger) occur if the null hypothesis were true? We simulate each frequency interval separately, because each has a different probability of success:

```matlab
>> for m = 1:6
    xa_(m, :) = binornd(na, p0(m), 1, 100000);
    xb_(m, :) = binornd(nb, p0(m), 1, 100000);
end
```

Next, we calculate the mean difference in the proportions correct for the two simulated conditions:

```matlab
>> pa_ = xa_ / na;
>> pb_ = xb_ / nb;
>> pd_ = pa_ - pb_;
>> mpd_ = mean(pd_);
```

Finally, we count the number of times that the simulated mean difference was equal to or greater than the observed mean difference, and divide by the number of simulations to obtain the probability estimate:

```matlab
>> extreme = mpd_ >= mpd;
>> p_sig = sum(extreme) / 100000
p_sig =
 0.0382
```

Thus, according to this paired comparison using binomial statistics, the SPP condition has significantly better scores than the MPPU condition. The execution time was about 3 seconds. This method was implemented by the author in a MATLAB function, which is called as:

```matlab
Difference_between_paired_proportions([xa,xb],[na,nb],'monte carlo');
```

### 16.5 Psychometric functions

A psychometric function describes the dependence of a subject's performance on some physical attribute of the stimulus. The subject performs a number of trials $n$ at a set of stimulus variables $u$, giving a vector of number of successes $x$. We then model the probability of success as a smooth function of the stimulus variable $u$, characterised by a set of parameters $\theta$:

$$ p = \Psi(u, \theta) $$

(16.11)

i.e. for each hypothetical set of parameters $\theta$, we can determine the probability of success $p$ at each value of $u$. The likelihood function $L(\theta)$ is the probability of obtaining the entire data set $x$,
given a hypothetical parameter set \( \theta \). A maximum-likelihood procedure searches through the parameter space to find the set \( \theta_m \) that gives the largest value of \( L \).

The psychometric functions in this thesis were fitted using the psignifit toolbox version 2.5.6 for MATLAB by Jeremy Hill (available at http://bootstrap-software.org/psignifit/) which implements the maximum-likelihood method described by Wichmann and Hill (2001). A maximum-likelihood procedure will find the parameters that generate the best fit of a specified model to the observed scores, however, it is still important to assess how good the fit is. Wichmann and Hill (2001) define a measure called deviance to assess the fit. Deviance is analogous to the sum of the squared differences between the model predictions and the subject's scores, thus a better fit yields a smaller deviance.

Because it is summed over all data points, deviance generally increases as the number of data points being fitted rises. In §8.5.4, deviance was used to compare alternative models, to judge which best fits the experimental data. Each subject had performed a different number of trials, thus the range of deviance varied across subjects. So that all subjects could be shown on a single graph (Figure 8.19), the deviances were normalised by dividing by the number of data points used in the fit. This normalisation does not affect the comparison between models within each subject.

### 16.6 Comparing conditions having different parameter sets

Section §16.4 described the comparison of two conditions with a repeated-measures design, where each condition was tested with the same set of stimulus parameter values. However, in some experimental designs, the stimulus parameter values are chosen adaptively, based on the performance of the subject up to that point (§2.3.3). The underlying assumption is that performance is characterised by a psychometric function. It is then likely that the two conditions to be compared will have a different set of stimulus parameter values.

One approach is to use apply a repeated-measures analysis method to the common subset of stimulus parameter values. The disadvantage is that this often omits regions of the stimulus parameter space where there is large difference in performance. A better approach is to calculate a summary statistic for each condition. This statistic should be an overall measure of performance. The statistic for each condition can then be compared.

If different stimulus parameters were used for different subjects and conditions, a simple mean score across parameters is inappropriate. If a psychometric function can be fitted to the results for each condition, the model parameters (e.g. the threshold) can be compared. However, in the Modified Melodies results for cochlear implant recipients (§11.5) it was apparent that a smooth psychometric function would be a poor fit. Instead, the following approach was taken.
16.6.1 Modified Melodies summary score

The summary score was calculated as the area under the percent correct graph, as illustrated in Figure 16.6. The top panel shows the Modified Melodies Stretch scores for subject M02 for the CI and Bimodal conditions (taken from §11.5, Figure 11.23). The middle panel shows the area under the graph for the Bimodal condition, and the bottom panel shows the area for the CI condition. It is clear that the area is larger for the Bimodal condition, indicating better performance. Note that the Bimodal condition was tested at more difficult stretch values (closer to 1.0), and the high score for stretch 1.5 contributes to the better summary score. Both graphs are interpolated to a chance score for stretch factor 1.0 (in the centre of the graph) which is inherent in the design of the test (§11.2.2). The CI condition was not tested at stretch 1.5 because chance performance had already been reached at stretch 2.0; effectively the CI score at stretch 1.5 is estimated by linear interpolation. This method allows a comparison between the two conditions that uses all available data points, unlike a paired comparison using the common subset of stretch values.

Figure 16.6 Modified Melodies summary scores for subject M02
16.6.2 Modified Melodies group median

A similar issue arises when attempting to find a measure of group performance when subjects were tested at different parameter values. Continuing with the example of the Modified Melodies contralateral hearing aid study (§11.5), stretch factors progressively closer to 1.0 were tested until chance scores were obtained. For example, as mentioned above, subject M02 was not tested at stretch 1.5 in the CI condition. However, if subjects such as M02 were simply omitted when calculating the group median result for stretch 1.5, it would clearly bias the result, as only the better performing subjects would be included. Instead, the missing scores for such subjects were estimated by linear interpolation towards chance, as shown in Figure 16.6.
17 References


Advanced Bionics Corporation (2007). "HiRes with Fidelity 120 Clinical Results", Valencia, CA, USA.


References


References


Houtsma, AJM and Smurzynski, J (1990). "Pitch identification and discrimination for..."


References


References


Wilson, BS, Lawson, DT, Finley, CC and Zerbi, M (1991b). "Speech processors for auditory prostheses, NIH Contract N01-DC-9-2401, 10th Quarterly Progress


